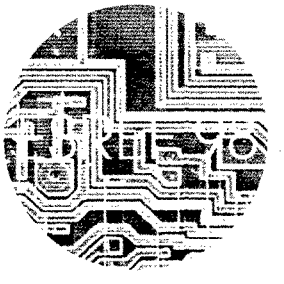
Proceedings of the
9th European Symposium on

# RELIABILITY OF ELECTRON DEVICES, FAILURE PHYSICS AND ANALYSIS



**5–9 October 1998**
**Copenhagen, Denmark**

*Edited by*

F. Jensen
C. Kjærgaard



PERGAMON

AQF99-02-0153

# REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

| 1. AGENCY USE ONLY (Leave blank) | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | 19 October 1998 | Conference Proceedings |

**4. TITLE AND SUBTITLE**

ESREF 98 - 9th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis

**5. FUNDING NUMBERS**

F61775-98-WE047

**6. AUTHOR(S)**

Conference Committee

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

Relaibility Consultancy APS
Pile Alle 11
Holte DK-2840
Denmark

**8. PERFORMING ORGANIZATION REPORT NUMBER**

N/A

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

EOARD
PSC 802 BOX 14
FPO 09499-0200

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

CSP 98-1017

**11. SUPPLEMENTARY NOTES**

**12a. DISTRIBUTION/AVAILABILITY STATEMENT**

Approved for public release; distribution is unlimited.

**12b. DISTRIBUTION CODE**

A

**13. ABSTRACT (Maximum 200 words)**

The Final Proceedings for ESREF 98 - 9th European Symposium on Reliability of Electron Devices, 5 October 1998 - 9 October 1998

This is an interdisciplinary conference. Topics include reliability of electron devices, including MEMs, failure physics and analysis.

**14. SUBJECT TERMS**

EOARD, Electronic Devices, Electronics and Electrical Engineering, Sensor Technology, Microelectronics, MEMs,

**15. NUMBER OF PAGES**

530

**16. PRICE CODE**

N/A

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| UNCLASSIFIED | UNCLASSIFIED | UNCLASSIFIED | UL |

# MICROELECTRONICS RELIABILITY

## SPECIAL ISSUE
## RELIABILITY OF ELECTRON DEVICES, FAILURE PHYSICS AND ANALYSIS


EDITORS-IN-CHIEF

**N.D. Stojadinović** **M.G. Pecht**

GUEST EDITORS

**F. Jensen** **C. Kjærgaard**



19981110 011

DTIC QUALITY INSPECTED 4





PERGAMON

## Aims and Scope

*MICROELECTRONICS RELIABILITY* is dedicated to disseminating the latest research results and related information on the reliability of microelectronic devices, circuits and systems. The coverage of the journal includes the following topics: physics and analysis; evaluation and prediction; modelling and simulation; methodologies and assurance. Papers which combine reliability with other important areas of microelectronic engineering circuits and systems, such as design, fabrication, packaging and testing, will also be welcome, and practical papers reporting case studies in the field are particularly encouraged.

Most accepted papers will be published as **Research Papers**, describing significant advances and completed work. Papers reviewing important developing topics of general interest may be accepted for publication as **Review Papers**. Urgent communications of a more preliminary nature and short reports on completed practical work of current interest may be considered for publication as **Research Notes**. All contributions are subject to peer review by leading experts in the field.

Additional regular features will include:

- special issues devoted to significant international conferences, or to important developing topics
- letters to the Editors
- industrial news and updates
- calendar of forthcoming events
- book reviews.

*MICROELECTRONICS RELIABILITY* is an indispensable forum for the exchange of knowledge and experience between microelectronics reliability professionals from both academic and industrial environments, and all those associated in any way with a steadily growing microelectronics industry and its many fields of application.

---

**© 1998 Elsevier Science Ltd. All rights reserved.**

This journal and the individual contributions contained in it are protected under copyright by Elsevier Science Ltd, and the following terms and conditions apply to their use:

**Photocopying**

Single photocopies of single articles may be made for personal use as allowed by national copyright laws. Permission of the publisher and payment of a fee is required for all other photocopying, including multiple or systematic copying, copying for advertising or promotional purposes, resale, and all forms of document delivery. Special rates are available for educational institutions that wish to make photocopies for non-profit educational classroom use.

Permissions may be sought directly from Elsevier Science Rights & Permissions Department, PO Box 800, Oxford OX5 1DX, UK; phone: (+44) 1865 843830, fax: (+44) 1865 853333, e-mail: permissions@elsevier.co.uk. You may also contact Rights & Permissions directly through Elsevier's home page (http://www.elsevier.nl), selecting first 'Customer Support', then 'General Information', then 'Permissions Query Form'.

In the USA, users may clear permissions and make payments through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA; phone: (978) 7508400, fax: (978) 7504744, and in the UK through the Copyright Licensing Agency Rapid Clearance Service (CLARCS), 90 Tottenham Court Road, London W1P 0LP, UK; phone: (+44) 171 436 5931; fax: (+44) 171 436 3986. Other countries may have a local reprographic rights agency for payments.

**Derivative Works**

Subscribers may reproduce tables of contents or prepare lists of articles including abstracts for internal circulation within their institutions. Permission of the publisher is required for resale or distribution outside the institution.

Permission of the publisher is required for all other derivative works, including compilations and translations.

**Electronic Storage or Usage**

Permission of the publisher is required to store or use electronically any material contained in this journal, including any article or part of an article. Contact the publisher at the address indicated.

Except as outlined above, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior written permission of the publisher.

Address permissions requests to: Elsevier Science Rights & Permissions Department, at the mail, fax and e-mail addresses noted above.

**Notice**

No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made.

Although all advertising material is expected to conform to ethical (medical) standards, inclusion in this publication does not constitute a guarantee or endorsement of the quality or value of such product or of the claims made of it by its manufacturer.

♾ The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper)




**ContentsDirect**, the ***free*** e-mail alerting service, delivers the table of contents for this journal directly to your PC, prior to publication. The quickest way to register for **ContentsDirect** is via the Internet at: **http://www.elsevier.nl/locate/ContentsDirect**. If you don't have access to the Internet you can register for this service by sending an e-mail message to cdsubs@elsevier.co.uk —specifying the title of the publication you wish to register for.

# SPECIAL ISSUE
# 9th European Symposium on RELIABILITY OF ELECTRON DEVICES, FAILURE PHYSICS AND ANALYSIS (ESREF 98)


Editorial ix
C. KJÆRGAARD, F. JENSEN

**OPENING SESSION**

Full-chip reliability analysis 851
D. OVERHAUSER, J. R. LLOYD, S. ROCHEL, G. STEELE, S. Z. HUSSAIN

An analysis of the quality and reliability supplement to the SIA roadmap 861
R. W. THOMAS

**THEME 1: FAILURE ANALYSIS AND CHARACTERIZATION**

The use of the focused ion beam in failure analysis 869
D. VERKLEIJ

Analysis of $I_{ddq}$ failures by spectral photon emission microscopy 877
M. RASRAS, I. DE WOLF, H. BENDER, G. GROESENEKEN, H. E. MAES, S. VANHAEVERBEKE, P. DE PAUW

Near-field-optical-probe induced resistance-change-detection (NF-OBIRCH) method for identifying defects in Al and TiSi interconnects 883
K. NIKAWA, T. SAIKI, S. INOUE, M. OHTSU

A new adaptive amplifier for biased electron beam induced current applications 889
M. CIAPPA, P. MALBERTI, P. FURCAS, M. VANZI

Detailed investigation of SEM-results by TEM at one sample using FIB-technique 895
U. MÜHLE, A. WIESNER, S. SCHRAY

Induced damages on CMOS and bipolar integrated structures under focused ion beam irradiation 901
J. BENBRIK, P. PERDU, B. BENTEO, R. DESPLATS, N. LABAT, A. TOUBOUL, Y. DANTO

Junction delineation and EBIC on FIB cross section 907
G. PEREZ, F. COURTADE, B. BENTEO, J-L. GAUFFIER, JL. KWANG

Direct observation of local strain field for ULSI devices 913
N. HASHIKAWA, K. FUKUMOTO, T. KUROI, M. IKENO, Y. MASHIKO

A new experimental technique to evaluate the plasma induced damage at wafer level testing 919
L. PANTISANO, A. PACCAGNELLA, L. PETTARIN, A. SCARPA, G. VALENTINI, L. BALDI, S. ALBA

Low frequency noise analysis as a diagnostic tool to assess the quality of 0.25$\mu$m Ti-silicided poly lines 925
E. P. VANDAMME, I. DE WOLF, A. LAUWERS, L. K. J. VANDAMME

Hot carrier degradation mechanisms in sub-micron *p* channel MOSFETs: impact on low frequency (1/f) noise behaviour 931
E. SHEEHAN, P. K. HURLEY, A. MATHEWSON

Reliability of nitrided wet silicon dioxide thin films in WSi2 or TaSi2 polycide process: influence of the nitridation temperature 937
K. YCKACHE, P. BOIVIN, F. BAIGET, S. RADJAA, G. AURIEL, B. SAGNES, J. OUALID, A. GLACHANT

**THEME 2: EOBT—ELECTRON AND OPTICAL BEAM TESTING**

Assembly and analysis of quantum devices using SPM based methods 943
L. MONTELIUS, T. JUNNO, S.-B. CARLSSON, M. H. MAGNUSSON, K. DEPPERT, H. XU, L. SAMUELSON

Circuit internal logic analysis with electric force microscope- (EFM-) testing 951
J. BANGERT, E. KUBALEK

Failure analysis of integrated devices by scanning thermal microscopy (SThM) 957
G. B. M. FIEGE, V. FEIGE, J. C. H. PHANG, M. MAYWALD, S. GÖRLICH, L. J. BALK

Scanning near-field optical microscopy analyses of electronic devices 963
R. M. CRAMER, W. R. SCHADE, R. HEIDERHOFF, L. J. BALK, R. CHIN

A new test method for contactless quantitative current measurement via scanning magneto-resistive probe microscopy 969
S. BAE, A. SCHLENSOG, W. MERTIN, E. KUBALEK, M. MAYWALD

Automatic fault tracing by successive circuit extraction from CAD layout data with the CAD-linked EB test system 975
K. MIURA, K. NAKAMAE, H. FUJIOKA

Cantilever influence suppression of contactless IC-testing by electric force microscopy 981
V. WITTPAHL, U. BEHNKE, B. WAND, W. MERTIN

Application of layout overlay for failure analysis 987
C. BURMER, S. GÖRLICH, S. PAUTHNER

Failure analysis of wafer using backside OBIC method 993
S. ITO, H. MONMA

**THEME 3: FAILURE MECHANISMS AND MODELLING**

ESD protection methodology for deep-sub-micron CMOS 997
K. BOCK, G. GROESENEKEN, H. E. MAES

Overview of the kinetics of the early stages of electromigration under low (= realistic) current density stress 1009
J. VAN OLMEN, J. V. MANCA, W. DE CEUNINCK, L. DE SCHEPPER, V. D'HAEGER, A. WITVROUW, K. MAEX

Effects of alloying elements on electromigration 1015
R. SPOLENAK, O. KRAFT, E. ARZT

A comparison between normally and highly accelerated electromigration tests 1021
S. FOLEY, A. SCORZONI, R. BALBONI, M. IMPRONTA, I. DE MUNARI,
A. MATHEWSON, F. FANTINI

Electromigration failure modes in damascene copper interconnects 1029
L. ARNAUD, R. GONELLA, G. TARTAVEL, J. TORRÈS, C. GOUNELLE, Y. GOBIL,
Y. MORAND

The dependence of stress induced voiding on line width studied by conventional and high resolution resistance measurements 1035
A. WITVROUW, K. MAEX, W. DE CEUNINCK, G. LEKENS, J. D'HAEN,
L. DE SCHEPPER

Lateral interface effect on pulsed DC electromigration analysis 1041
P. WALTZ, L. ARNAUD, G. LORMAND, G. TARTAVEL

Mechanical stress evolution and the Blech length: 2D simulation of early electromigration effects 1047
V. PETRESCU, A. J. MOUTHAAN, W. SCHOENMAKER, C. SALM

Systematic derivation of latchup design rules for submicron CMOS processes from test structures 1051
J. A. VAN DER POL, P. B. M. WOLBERT

Reversibility of charge trapping and SILC creation in thin oxides after stress/anneal cycling 1057
P. RIESS, R. KIES, G. GHIBAUDO, G. PANANAKAKIS, J. BRINI

Precise quantitative evaluation of the hot-carrier induced drain series resistance degradation in LATID-n-MOSFETs 1063
G. H. WALTER, W. WEBER, R. BREDERLOW, R. JURK, C. G. LINNENBANK,
C. SCHLÜNDER, D. SCHMITT-LANDSIEDEL, R. THEWES

On-wafer heating tests to study stability of silicon devices 1069
D. MANIC, Y. HADDAB, R. S. POPOVIC

Characterization of SILC in thin-oxides by using MOSFET substrate current 1075
B. DE SALVO, G. GHIBAUDO, G. PANANAKAKIS, F. MONDON

Hot carrier induced device degradation in RF-nMOSFET's 1081
J. T. PARK, B. J. LEE, D. W. KIM, C. G. YU, H. K. YU

The effect of hot electron current density on nMOSFET reliability 1085
O. BUIU, S. TAYLOR, I. S. AL-KOFAHI, C. BEECH

Dependence of gate oxide breakdown on initial charge trapping under Fowler-Nordheim injection 1091
A. MARTIN, R. DUANE, P. O'SULLIVAN, A. MATHEWSON

Modelling and simulation of hot-carriers degradation of high voltage floating lateral NDMOS transistors 1097
E. VANDENBOSSCHE, C. DE KEUKELEIRE, M. DE WOLF, H. VAN HOVE, J. WITTERS

A new hot carrier degradation law for MOSFET lifetime prediction 1103
B. MARCHAND, G. GHIBAUDO, F. BALESTRA, G. GUÉGAN, S. DELEONIBUS

Recovery and stress dynamics in bipolar transistors and MOS devices 1109
F. INGVARSON, L-Å. RAGNARSSON, P. LUNDGREN

Electrical parameters degradation law of MOSFET during ageing 1115
CH. MOURRAIN, CH. TOURNIOL, M. J. BOUZID

Investigation of the intrinsic $SiO_2$ area dependence using TDDB testing and model integration into the design process 1121
J. PRENDERGAST, N. FOLEY, J. SUEHLE

Two-step stress method for the dynamic testing of very thin (8 nm) $SiO_2$ films 1127
R. RODRÍGUEZ, E. MIRANDA, M. NAFRÍA, J. SUÑÉ, X. AYMERICH

Temperature dependence of snap back breakdown up to 300°C analyzed using circuit level model and simulation 1133
D. UFFMANN

Modelling the field soft error rate of DRAMs by varying the critical cell charge 1139
H. SCHLEIFER, O. KOWARIK, K. HOFFMANN, W. RECZEK

Design of a low EMI susceptibility CMOS transimpedance operational amplifier 1143
G. SETTI, N. SPECIALE

Gate bias stress in hydrogenated and unhydrogenated polysilicon thin film transistors 1149
B. TALA-IGHIL, H. TOUTAH, A. RAHAL, K. MOURGUES, L. PICHON, F. RAOULT, O. BONNAUD, T. MOHAMMED-BRAHIM

**THEME 4: FAILURE PREVENTION STRATEGIES**

Use of wafer maps in integrated circuit manufacturing 1155
C. K. HANSEN, P. THYREGOD

Field failure analysis on transmission data equipment due to lightning discharges 1165
E. MINO DIAZ, J. E. VILA ARESTÉ

Comparison between field reliability and new prediction methodology on avionics embedded electronics 1171
P. CHARPENEL, P. CAVERNES, V. CASANOVAS, J. BOROWSKI, J. M. CHOPIN

Improved reliability of bistable circuits by selective hot-carrier stress reduction 1177
A. G. M. DAS, S. JOHNSON

A study of NMOS behavior under ESD stress: simulation and characterization 1183
A. Z. WANG, C. TSAY, A. LELE, P. DEANE

The time of "guessing" your failure time distribution is over! 1187
K. CROES, J. V. MANCA, W. DE CEUNINCK, L. DE SCHEPPER, G. MOLENBERGHS

Extended noise analysis - a novel tool for reliability screening 1193
G. HÄRTLER, U. GOLZE, K. PASCHKE

## THEME 5: COMPOUND SEMICONDUCTOR AND OPTOELECTRONIC RELIABILITY

Study of degradation mechanisms in compound semiconductor based devices by SEM cathodoluminescence 1199
G. SALVIATI, C. ZANOTTI-FREGONARA, M. BORGARINO, L. LAZZARINI, L. CATTANI, P. COVA, M. MAZZER

Degradation behavior in InGaAs/GaAs strained-quantum well lasers 1211
T. TAKESHITA, M. SUGO, T. NISHIYA, R. IGA, M. FUKUDA, Y. ITAYA

Early signatures for REDR-based laser degradations 1215
A. BONFIGLIO, M. B. CASU, M. VANZI, R. DE PALO, F. MAGISTRALI, G. SALMINI

Coupling technology impact on low-cost laser modules performances and reliability 1221
M. MORIN, J.-P. DEFARS, P. DEVOLDÈRE

Failure mechanisms of Schottky gate contact degradation and deep traps creation in AlGaAs/InGaAs PM-HEMTs submitted to accelerated life tests 1227
G. MENEGHESSO, C. CROSATO, F. GARAT, G. MARTINES, A. PACCAGNELLA, E. ZANONI

Pulsed current stress of berillium doped AlGaAs/GaAs HBTs 1233
L. CATTANI, M. BORGARINO, F. FANTINI

Degradation of performance in MESFETs and HEMTs: simulation and measurement of reliability 1239
T. FENG, N. STRIFAS, A. CHRISTOU

A new method for temperature mapping on GaAs field effect transistors 1245
E. MARTIN, J. P. LANDESMAN, P. BRAUN, A. FILY

## THEME 6: MICROSYSTEMS RELIABILITY

Reliability of industrial packaging for microsystems 1251
R. DE REUS, C. CHRISTENSEN, S. WEICHEL, S. BOUWSTRA, J. JANTING, G. FRIIS ERIKSEN, K. DYRBYE, T. ROMEDAHL BROWN, J. P. KROG, O. SØNDERGÅRD JENSEN, P. GRAVESEN

Analysis of thermomechanical stresses in a 3D packaged micro electro mechanical system 1261
C. PELLET, M. LECOUVÉ, H. FRÉMONT, A. VAL, D. ESTÈVE

Electrical characterization and modification of a microelectromechanical system (MEMS) for extended mechanical reliability and fatigue testing 1265
D. MEUNIER, R. DESPLATS, J. BENBRIK, G. PEREZ, C. PELLET, D. ESTÈVE, B. BENTEO

Structures for piezoresistive measurement of package induced stress in transfer molded silicon pressure sensors 1271
J. B. NYSÆTHER, A. LARSEN, B. LIVERØD, P. OHLCKERS

## THEME 7: PACKAGING

Electronic systems packaging: future reliability challenges 1277
J. BARRETT

A concept to relate wire bonding parameters to bondability and ball bond reliability 1287
Z. N. LIANG, F. G. KUPER, M. S. CHEN

Measurement of the thermomechanical behaviour of the solder-lead interface in solder joints by laser probing: a new method for measuring the bond quality 1293
S. DILHAIRE, A. CORNET, E. SCHAUB, C. RAUZAN, W. CLAEYS

Experimental design and evaluation of interconnection materials for improvement of joint reliability at power transistors 1297
P. JANSSON

Crack mechanism in wire bonding joints 1301
S. RAMMINGER, P. TÜRKES, G. WACHUTKA

Materials interfaces in flip chip interconnects for optical components; performance and degradation mechanisms 1307
R. H. ESSER, A. DIMOULAS, N. STRIFAS, A. CHRISTOU, N. PAPANICOLAU

A study of soldering heat evaluation for SMDs 1313
Y. ETOH, T. KAYAMA, K. SASAKI

**THEME 8: POWER DEVICES**

Advanced IGBT modules for railway traction applications: reliability testing 1319
H. BERG, E. WOLFGANG

Some observation dealing with the failures of IGBT transistors in high power converters 1325
S. JANUSZEWSKI, M. KOCISZEWSKA-SZCZERBIK, H. ŚWIĄTEK

Strain depending reliability of automotive diodes 1331
L. GALATEANU, M. G. STOICA, E. POPA

Extrapolation of cosmic ray induced failures from test to field conditions for IGBT modules 1335
C. FINDEISEN, E. HERR, M. SCHENKEL, R. SCHLEGEL, H. R. ZELLER

Thermomechanical deformation imaging of power devices by electronic speckle pattern interferometry (ESPI) 1341
K. NASSIM, L. JOANNES, A. CORNET, S. DILHAIRE, E. SCHAUB, W. CLAEYS

On the effect of power cycling stress on IGBT modules 1347
P. COVA, F. FANTINI

Temperature measurements and thermal modeling of high power IGBT multichip modules for reliability investigations in traction applications 1353
A. HAMIDI, G. COQUERY, R. LALLEMAND, P. VALES, J. M. DORKEL

On-chip reliability investigations on power modules actually working in inverter systems 1361
T. FRANKE, M. HONSBERG-RIEDL, P. SIMON, J. OTTO, S. RAMMINGER, G. SOELKNER, E. WOLFGANG

Author Index I

PERGAMON

Microelectronics Reliability 38 (1998) ix

MICROELECTRONICS RELIABILITY

# Editorial

This Special Issue of *Microelectronics Reliability* is devoted to papers presented at the European Symposium on Reliability of Electron Devices, Failure Physics and Analysis (ESREF), which is the annual forum for reliability physics and analysis of electronic components in Europe.

The 1998 Copenhagen event is the 9th conference in the series. In 1991, ESREF merged with the International Conference on Quality in Electronic Components, Failure Prevention, Detection and Analysis. It now merges with the European EOBT Conference on Electrical and Optical Beam Testing. The ESREF conferences are seen to target a very broad cross-section of research in the fields of electronic component reliability and failure analysis, and each year an increasing numbers of papers and participants from all over the world contribute to making this a truly international event.

Whilst the major emphasis is on silicon integrated circuits, there are also sessions devoted to compound semiconductor reliability and optoelectronic devices, to packaging techniques and, for the first time, a session is given to a new major growth area, MEMS, or microelectromechanical systems.

This year, more than 70 papers, selected from more than 100 abstracts received, will be presented at the conference. Additionally, each of the eight themes will be introduced by an invited paper presented by an internationally recognised expert. Furthermore, two significant invited papers introduce the technical sessions at the start of the conference.

In a new tradition, the best IRPS98 paper (International Reliability Physics Symposium, USA) was presented at the conference, as was the best paper from the Japanese RCJ Reliability Symposium, 1997. This follows an exchange agreement between ESREF and these two international conferences in the USA and Japan. Copyright restrictions preclude the inclusion of the former in this issue.

The Editors of this Special Issue wish to thank the session chairmen of the Technical Programme Committee for the enormous amount of work and positive collaboration they put into the selection and organization of the papers for the conference sessions.

The Editors also thank all authors for their contributions to the conference and to these proceedings.

Finally, we appreciate the guidance on putting these proceedings together given to us by the Publisher.

Claus Kjærgaard
Finn Jensen
Guest Editors

0026-2714/98/$ – see front matter. © 1998 Published by Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00167-X

# ESREF STEERING COMMITTEE

| | |
|---|---|
| L. J. Balk | Bergische Universität, Germany |
| M. Barré | MATRA Defense, France |
| G. M. Brydon | QaRel, United Kingdom |
| Y. Danto | IXL, University of Bordeaux, France |
| F. Fantini | University of Modena, Italy |
| W. Gerling | Siemens AG, Germany |
| Ö. Hallberg | Ericsson Telecom AB, Sweden |
| H. E. Maes | IMEC, Belgium |
| A. J. Mouthaan | University of Twente, The Netherlands |

# ESREF 98 ORGANIZING COMMITTEE

| | |
|---|---|
| Finn Jensen | Conference Chairman<br>FJRC, Denmark |
| Jørgen Møltoft | Conference Vice Chairman<br>IAE/DTU, Denmark |
| Claus Kjærgaard | Technical Programme Chairman<br>IAE/DTU, Denmark |
| Örjan Hallberg | Technical Programme Vice Chairman<br>Ericsson Telecom, Sweden |
| Lars Rimestad | Tutorial Chairman<br>IAE/DTU, Denmark |

# ESREF 98 TECHNICAL PROGRAMME COMMITTEE

## Chairmen and co-chairmen

| | |
|---|---|
| J.-P. Fortea | CNES, France |
| M. Vanzi | University of Cagliari, Italy |
| L. J. Balk | Bergische Universität, Germany |
| W. Claeys | Université de Bordeaux, France |
| G. Ghibaudo | LPCS, ENSERG, France |
| A. J. Mouthaan | University of Twente, The Netherlands |
| A. C. Brombacher | TUE, The Netherlands |
| D. Schmitt-Landsiedel | Technische Universität München, Germany |
| F. Fantini | University of Modena, Italy |

| | |
|---|---|
| G. M. Brydon | QaRel, United Kingdom |
| S. Bouwstra | Mikroelektronik Centeret, Denmark |
| J. Branebjerg | DELTA, Denmark |
| C. Olsson | Ericsson, Sweden |
| J. Lloyd | Lloyd Technology Associates, USA |
| E. Wolfgang | Siemens, Germany |
| M. Ciappa | ETH, Switzerland |

## Members

| | |
|---|---|
| M. D. Trias | SERMA Technologies, France |
| O. Bonnaud | University of Rennes, France |
| D. S. C. Chan | National University of Singapore, Singapore |
| E. Demm | Siemens AG, Germany |
| B. K. Jones | Lancaster University, UK |
| A. Konczakowska | Technical University of Gdansk, Poland |
| H. Koyama | Mitsubishi Electric Corporation, Japan |
| K. Tenenbaum | Ericsson Components AB, Sweden |
| L. K. J. Vandamme | Eindhoven University of Technology, The Netherlands |
| W. Claassen | Philips Semiconductors, The Netherlands |
| K. M. A. Van Doorselaer | Bell Alcatel, Belgium |
| J. C. Clement | Thomson TMM, France |
| W. Mertin | Gerhard Mercator Universität, Germany |
| L. Montelius | University of Lund, Sweden |
| H. Fujioka | Osaka University, Japan |
| S. Görlich | Siemens AG, Germany |
| E. Gornik | Institut für Festkörperelektronik, Germany |
| P. Koschinski | Ams-Topometrix, Germany |
| I. De Wolf | IMEC, Belgium |
| A. Majumdar | University of California, USA |
| B. Curtois | Tima, France |
| J. Phang | National University of Singapore, Singapore |
| X. Aymerich | Universitat Autonoma Barcelona, Spain |
| S. Cristoloveanu | ENSERG, France |
| H. Gieser | Fhg/IFT, Germany |
| C. D. Graas | Texas Instruments, USA |
| G. Groeseneken | IMEC, Belgium |
| P. Mortini | SGS Thomson Microelectronics, France |
| N. Stojadinovic | University of Nis, Yugoslavia |
| A. Touboul | University of Bordeaux, France |
| L. De Schepper | Limburgs Universitair Centrum, Belgium |
| A. Martin | NMRC, Ireland |

| | |
|---|---|
| A. Scorzoni | CNR-LAMEL, Italy |
| W. Gerling | Siemens A.G, Germany |
| J. A. Van Der Pol | Philips Semiconductor, The Netherlands |
| R. W. Thomas | Technology Experts Network, USA |
| C. Canali | University of Padova, Italy |
| J. M. Dumas | ENSIL, France |
| M. Fukuda | NTT Optoelectronics Laboratories, Japan |
| A. A. Rezazadeh | King's College London, UK |
| C. Thome | Dassault Elektronique, France |
| A. Christou | University of Maryland, USA |
| J. Barrett | NMRC, Ireland |
| Y. Danto | University of Bordeaux, France |
| A. Amerasekara | Texas Instruments, USA |
| N. Sinnadurai | Middlesex University, UK |
| B. Eriksson | Ericsson AB, Sweden |
| E. Zanoni | University of Padova, Italy |
| H. L. Hartnagel | TH Darmstadt, Germany |
| F. Magistrali | Pirelli Cavi, Italy |
| S. Ramminger | Siemens Corporate Technology, Germany |
| D. Reznik | Siemens Corporate Technology, Germany |
| G. Sölkner | Siemens Corporate Technology, Germany |
| G. Deboy | Siemens Corporate Technology, Germany |
| P. Türkes | Siemens Corporate Technology, Germany |
| T. Franke | Siemens Corporate Technology, Germany |
| C. Zardini | University of Bordeaux, France |
| H. Kabza | University Ulm, Germany |
| W. Mertin | University Duisburg, Germany |
| P. Voss | EUPEC, Germany |
| H.-R. Zeller | ABB Semiconductors, Switzerland |
| E. Herr | ABB Semiconductors, Switzerland |
| R. Zehringer | ABB Research, Switzerland |
| D. Newcombe | Mitel Semiconductors, UK |
| M. Rahimo | Mitel Semiconductors, UK |
| M. Ciappa | ETH Zürich, Switzerland |
| G. Coquery | INRETS, France |

# SPONSORS AND SUPPORT

We wish to thank the following for their contribution to the success of this conference:

United States Air Force European Office of Aerospace Research and Development
TeleDanmark
Ericsson Telecom AB
The Technical University of Denmark

We are happy also to acknowledge the technical co-sponsorship of:

IEEE Electron Devices Society, USA
VDE/ITG - Germany
SRE - Scandinavian Reliability Engineers

# EQUIPMENT EXHIBITORS and SERVICE PROVIDERS

AETRIUM Inc.
CARL ZEISS JENA GmbH
DELTA
DESTIN N. V.
ELSEVIER
HAMAMATSU PHOTONICS
HTT High Tech Trade GmbH
John Wiley & Sons
MASER Engineering B.V.
MICROTRON ADVANCED TECHNOLOGY
SERMA TECHNOLOGIES
SIMAC MASIC
SONOSCAN
SYNATRON GmbH
QUALITAU Ltd
QUANTUM FOCUS INSTRUMENTS Corp.
LCIE
NEXTRAL

# BEST PAPER AWARDS ESREF 1997

A reliability study of titanium silicide lines using micro-Raman spectroscopy and emission microscopy
I. DE WOLF, D. J. HOWARD, M. RASRAS, A. LAUWERS, K. MAEX,
G. GROESENEKEN, H. E. MAES

## Best Student Paper Award

Thermal simulation and characterisation of the reliability of Thz Schottky diodes
M. BRANDT, M. SCHÜßLER, E. PARMEGGIANI, C. LIN, A. SIMON,
H. L. HARTNAGEL

## Best Poster Paper Award

Experimental analysis and 2D simulation of AlGaAs/GaAs HBT base leakage current
C. MANEUX, N. LABAT, N. SAYSSET, A. TOUBOUL, Y. DANTO

# ESREF 1998 SPECIAL INVITED PAPERS

## 1998 International Reliability Physics Symposium Best Paper Award

Disturbed bonding states in $SiO_2$ thin-films and their impact on time-dependent dielectric breakdown
J. McPHERSON, H. C. MOGUL

## 1997 Reliability Symposium on Electronic Devices, Japan Best Paper Award

A Study of soldering heat evaluation for SMDs
Y. ETOH, T. KAYAMA, K. SASAKI

PERGAMON

Microelectronics Reliability 38 (1998) 851–859

MICROELECTRONICS RELIABILITY

# Full-Chip Reliability Analysis

David Overhauser, J.R. Lloyd*, Steffen Rochel, Gregory Steele, and Syed Zakir Hussain

*Simplex Solutions, Inc.*
*521 Almanor, Sunnyvale, CA 94086*
**Lloyd Technology Associates, Inc.*
*Box 194, Stow MA 01775-0194*

**Abstract**

Reliability analysis has not been promoted to the realm of full-chip because techniques to extract, manage, and process full-chip power grid and signal data have not been previously available. This paper introduces techniques that have been developed to permit both full-chip power grid and signal net electromigration and Joule heating analysis. Results of this analysis provide feedback to the designer to permit easy design modification to provide superior "designed-in" long-term reliability.
© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Much research has been performed in order to understand the physics of device and interconnect reliability. Models and methodologies, such as those in BERT [1], have been developed to estimate the performance and reliability of small test circuits. However, due to the computational complexity of existing methodologies, these techniques have not been elevated to the realm of full-chip power grid and signal net reliability analysis.

Full-chip reliability analysis looks at reliability in the complete physical layout context and requires techniques to extract, manage, and process full-chip data. This has become more critical as advances in technology allow narrower interconnect structures and higher frequency designs. This combination increases the risk of electromigration and Joule heating induced failures unless they can be reduced or eliminated in the design process.

In order to insure that long term reliability is designed in, two challenging tasks must be performed. First, the reliability of the entire chip with all its circuit elements must be estimated in a realistic manner, and this information must be conveyed to the designer in a user-friendly way such that changes can be made.

Unfortunately, modern designs include tens of millions of resistors in a single power grid and may include hundreds of thousands of nets each containing many elements. Therefore, reliability analysis on designs of this size requires new techniques for circuit extraction, data management, and data analysis. Most existing extraction tools perform data reduction during extraction. This reduction is a mathematical wire macromodeling in which a small RC wire model is generated for a large distributed RC network of metal wires. This macromodel exhibits nearly the same response as the large original distributed RC network in SPICE simulators. Unfortunately, this macromodel has no physical parameters, such as wire width and length, so analysis techniques that are dependent on metal segment physical characteristics cannot perform analysis on reduced macromodels of chip wires. In addition, many layout extractors also cluster contacts, which means that a cluster of contacts connecting two metal layers is modeled as a single contact with a larger size. This can be described as a

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00152-8

macromodel for a group of contacts. This contact clustering, if performed during extraction, prevents analysis tools from modeling the impact of current crowding through contact arrays.

Traditionally, designers are given simple wire current density limits to which they must adhere. These limits are based on "worst case" estimates of the current density that is expected under use conditions, usually maximum temperature. Clearly this is overly restrictive. A simple calculation can show this assumption is overly conservative. If the industry-typical maximum current density of $2x10^5$ A/cm$^2$ were passed through most conductors, a chip dissipating kilowatts would be required. A kilowatt chip is, of course, patently ridiculous.

The greatest effect of using this overly conservative approach is that chip performance suffers greatly. Typically, in a chip, even a state of the art microprocessor, the vast majority of the conductor lines are passing little if any current during normal operation. If the real level of stress at use conditions can be factored into the reliability calculation, local violations of the design rules as commonly practiced in the design community today can be tolerated without compromising overall lifetime. This procedure can be called "Reliability Budgeting" and is used in some way in most high-performance chip design processes.

This methodology must analyze the circuit to obtain realistic estimates of actual currents flowing in the circuit, apply electromigration modeling, such as Black's equation, to individual wire segments and perform statistical analysis over the wires in the design to estimate the probability of failure over its intended lifetime. This cannot be accomplished without a detailed analysis of the entire chip.

The primary reason that full-chip electromigration analysis is necessary is that with growing chip complexities, designers no longer understand exactly how their chip operates. Due to complex power grids and distributed blocks in a design, current flow from a pin to the transistors cannot be determined without full-chip analysis. Adjacent blocks and their internal power routing have a significant impact on current distribution. In addition, since design tools do not consider electromigration limits, signal lines can experience excessive currents due to logic hazards and higher chip frequencies. Higher currents also increase the likelihood for signal line Joule heating failures.

The analysis of the full chip entails two separate challenges. Both signal and power grid analyses need to be considered. Signal analysis will be required for advanced technologies because a signal line failure is catastrophic to a design (an open line ceases functionality), whereas a power grid normally includes some redundancies.

This paper introduces several techniques that have been developed to perform full-chip electromigration and Joule heating analysis. A methodology has been developed to permit analysis of global chip power nets as well as signal nets. This methodology includes extraction of chip interconnect data without the application of interconnect net or power grid reduction or contact clustering, static and dynamic full-chip analysis to determine current loading characteristics at the various device contacts to interconnect, and modeling mechanisms to report either wire segments likely to fail or overall chip lifetime statistics. Static analysis is the analysis of the design without the use of test vectors, while dynamic analysis requires test vectors to analyze a design. The ability to apply reliability analysis to full-chip designs makes it possible to bring product reliability and reliability budgeting into the hands of designers.

## 2. Performing full-chip reliability analysis

Full-chip reliability analysis begins with the extraction from layout of the parasitic data associated with both signal and power networks. In order to support electromigration analysis as well as give meaningful feedback to designers, additional data is provided by the extractor, such as wire segment size, layout layer, and layout location. In addition, a transistor netlist is generated and the attachment of each transistor to each signal net is noted in the circuit database. The transistor network is then analyzed in either a static or dynamic (using test vectors) mode to obtain the characteristics of the current flow in the chip. This current flow data is then used to perform full-chip reliability analysis.

### *2.1. Reliability modeling*

Both power grid and signal electromigration analyses require the use of failure models. The choice of failure model must be made for each particular manufacturing process and design rules. Sub-micron lines will exhibit different failure kinetics than will wider lines and vias and contacts will exhibit yet other failure kinetics. Which kinetics to use must be determined from either testing or prior knowledge of similar processes.

Modeling is provided for each metal and via layer in the chip. Data from the process, in the form of median times to failure from experiments using appropriate test structures ($t_{50}$) and the deviation in the times to failure ($\sigma$) for wide and narrow metal lines must be provided as well as failure time data for vias and contacts.

Other important modeling parameters are based on the test structure design. In order to estimate the reliability, the chip must be represented as an aggregate of equivalent test structures. The current density in each equivalent test structure, called a "failure element" is determined and the probability of failure for each element is then calculated. For this analysis, each failure element is assumed to fail according to lognormal statistics [2]. The cumulative probability of failure at time, t, for a single failure element, $P_{fi}(t)$, is given approximately by [3]:

$$P_{fi}(t) = \frac{\sigma}{\sqrt{2\pi}\ln(t_{50}/t)} \exp - \left( \frac{\ln(t_{50}/t)}{\sqrt{2}\sigma} \right)^2, \quad (1)$$

where $\sigma$ is the lognormal standard deviation and $t_{50}$ is the median time to failure at use conditions calculated from the test structure data obtained at accelerated test conditions according to the generalized model:

$$t_{50} = AT^m j^{-n} \exp\left(\frac{\Delta H}{kT}\right), \quad (2)$$

where the temperature exponent, m, is either 0, 1, or 2 and the current exponent, n, is either 2 for nucleation dominated failure or 1 for growth dominated failure. In Eq. (2), $A$ is the technology and structure dependent pre-factor, $j$ is the current density, $\Delta H$ is the activation energy for the electromigration mechanism, $k$ is Boltzmann's constant, and $T$ is temperature. The selection of exponents m and n permit the application of a variety of failure models as follows:

| m | n | Failure model |
|---|---|---|
| 0 | 2 or measured | Generalized Black equation [4], n is 2 or experimentally determined |
| 2 | 2 | Nucleation dominated failure [5] |
| 1 | 2 | Stress nucleation dominated failure [6] |
| 1 | 1 | Growth dominated failure [7] |

After the failure probability for each element has been calculated using Eq. (1), the full chip reliability is then estimated from the extreme value lognormal distribution expressed as

$$P_f(n,t) = 1 - \prod_{i=1}^{n} \left(1 - P_{fi}(t)\right). \quad (3)$$

This is analogous to the weakest link in a chain, where each link is characterized by a lognormal failure distribution [2].

For each metal layer, the layer thickness and current density limits for peak, average, and RMS currents through the wire segments or failure elements are provided (different foundries provide different rules) for simple limit checks. These parameters are provided to analysis tools using model cards for each metal, contact, and via layer in the design. This is similar in usage to providing a device model card for each device type in a SPICE deck. A set of model parameters for the median time to failure (MTTF) is then developed. Different model parameters are applied for narrow wires and wide wires, where an additional model parameter defines the boundary between narrow and wide wires. Each via and contact model provides the current limits for peak, average, and RMS currents through each via/contact for simple limit checks, and a set of model parameters for the MTTF model. In this analysis, n=2 kinetics was used for conductor stripes and n=1 kinetics was used for vias and contacts.

The question of temperature is dealt with here in a simplified form. As long as the current density is kept below the critical level that leads to temperature gradient induced failure, the variations in temperature from point to point on a chip are considered unimportant. The temperature is used as an input parameter and this is assumed to be the metalization temperature throughout the chip. The user can estimate, from other means, the Joule heating that can be expected in the worst case, which can be used as an estimate of the metalization temperature throughout the chip. Since the conductors with the highest current density are also the hottest, these are the only ones that we need consider. If there are regions where the chip may be a few degrees cooler, the current density will also be lower. Since the failure probability is a sensitive function of the lifetime, these regions do not contribute to the overall failure probability and can

be ignored. Therefore it is an unnecessary complication to deal with the precise temperature profiles as long as temperature gradient induced failure is prevented by RMS current limit checks.

### 2.2. Power grid analysis

The power grid of a chip is operated primarily in a pulsed DC sense with respect to electromigration analysis. It has been shown that at operating frequencies and temperatures (~100°C and greater than ~10 kHz), electromigration driving force is determined by the average current density [8]. Thus, average current throughout the circuit is used to perform electromigration analysis on the grid. The full-chip transistor analysis tool [9] provides the average current drawn by each transistor connected to the power grid. Average current is defined as over 1 microsecond, a time well within the stochastic atomic jump frequency.

One at a time, each power grid is modeled by voltage sources at the pins providing power to the chip and the transistor tap currents at the device connection points. The large linear system is then solved to determine the precise current flowing through every wire segment and via in the chip [10]. Once each wire segment current density has been determined, simple checks are applied to identify those wires in the design which exceed a limit that had been predetermined from Eq. (3), making the worst case assumption that all circuit elements are operated at the limit current. A more detailed analysis of the reliability is then made by calculating the theoretical time to failure for each element that exceeds the limit and using the proper failure statistics to obtain a failure probability as a function of time for the entire chip. The results are highly dependent on the choice of the statistical model used. In this case, extreme value lognormal, also known as "multilognormal" statistics is used, where the failure probability is assumed to be that of a chain that fails at the weakest link. Any other failure distribution could be used, but experience and theoretical justifications suggest that the extreme value lognormal is appropriate. The wire segments with minimum MTTFs can be identified and this information can be provided to the designer for an engineering change order (ECO) if the overall chip probability is below specification.

This procedure can permit optimization of a design with respect to performance and reliability. Since the reliability of each circuit element can be determined, the reliability can be budgeted such that the overall reliability is satisfactory even if individual regions are exercised at current density levels that may be above those permitted in the traditional manner.

Another method to optimize the performance is to take advantage of the "Blech Length" effect [11]. When a metal conductor bounded by diffusion barriers such as contacts or W vias is subjected to electromigration producing conditions, a stress gradient is generated that produces a chemical potential gradient opposite to that of the electromigration force. While under electromigration conditions, the stress gradient will increase until a steady state is achieved, resulting in "Blech Condition" where the stress gradient and the electromigration force exactly balance. Since at a contact the mass flux vanishes, the Blech condition is produced throughout the conductor and electromigration comes to a complete halt.

The "Blech Product", jl, where j is the current density and l is the conductor length, characterizes this condition. If this product is below a critical level, electromigration will not take place.

For each current density, there will be a conductor length that will be able to satisfy the Blech Condition. Since we know the current flowing through each line, the conductors that are short enough can be identified. The situation is only slightly complicated for complex circuit elements with many branches, varying current density or varying width. However, this can be handled by a simple series of simultaneous equations to arrive at an effective length based on the current density distribution.

Great care must be exercised in the use of this technique, however. The Blech Product can be defined in terms of void nucleation or growth and would therefore depend on the failure criteria, i.e., whether a small void would be tolerated by the circuitry and by the amount of residual tensile stress in the metal film. Further discussion of this topic would be outside the scope of this paper, but in principle this can be taken into consideration allowing for higher than normal currents in shorter lines enabling an increase in performance.

Once analysis has been performed on a power grid feedback to the designer is provided in several forms: textual reports, GDS II overlays, and a graphical form. The graphical feedback of the power grid is color-coded, which allows for an easy identification of critical wire segments in the grid. The overall MTTF for the power grid is reported as well.

### 2.3. Signal line analysis

Signal electromigration analysis is more complex than power grid analysis. In addition to the electromigration lifetime based on the average current, the RMS current is also calculated. Joule heating depends on RMS current and it must be realized that RMS current always exceeds average current for any duty cycle less than one.

Joule heating produces temperature gradients that can cause failure due to temperature gradient induced flux divergences. In principle, the RMS current density must be limited to no more than that used in the accelerated tests from which the model parameters were obtained. In practice, this has usually been on the order of $2 \times 10^6$ A/cm$^2$ to obviate temperature gradient induced failure and this is what was chosen for this study.

RMS current is also treated differently. Instead of budgeting the reliability, the RMS current is treated as an absolute "speed limit" and exceeding the limiting value is not permitted. RMS current is also not included in the reliability calculation.

Due to the AC current behavior of signal nets in digital designs, signal wires must be checked for both average and RMS current density violations. Average checks find those wire segments with high levels of unidirectional current density impacting electromigration while RMS checks find those wire segments that suffer from Joule heating induced failure mechanisms. Nearly all signal lines include metal segments that exhibit DC current behavior while most metal segments exhibit AC current behavior. Signal electromigration analysis is performed on a net by net basis and requires the simulation of the charging and discharging of the signal net for all possible current paths in order to determine the worst case peak, average, and RMS current for each wire segment in the net. Computing all current parameters permits a wide range of electromigration checks and MTTF models to be applied.

Since detailed simulation of all nets on a chip is not practical, signal line analysis is split into two steps: a filtering step and a simulation step [12]. A set of novel filtering techniques has been developed to obviate the need for simulation of all nets. These filters are intended to easily identify those nets in a design, which are small in size or small in current, which will never experience electromigration or Joule heating induced failures. Both filters and simulation take into account:

- driver strength associated with each net,
- correlated distributed drivers on each net,
- actual parasitic and capacitive load attached to each net, and
- relative activity levels of the net.

Based on this information the peak, average and RMS current density in each net are calculated based on a static approach and checked against current density limits. As in power grid analysis, different limits are applied for narrow and wide wire segments.

The first filtering step grossly overestimates the current densities in each net's wire segments so that no net susceptible to failure will be missed. Nevertheless, our experience shows, that in a reasonably good design (from a reliability verification point of view), this filtering step reduces the number of nets to be considered further by at least an order of magnitude. All nets for which an overestimated current density is above any limit are marked for consideration in the subsequent more exact simulation step. Because the filtering step is completely static in nature, it can be applied on the full-chip level with hundreds of thousands of signal nets.

The next step is based on the simulation of the time dependent currents during charging and discharging phases of all the nets marked as critical after the filtering step. Each critical net is simulated for each possible charging and each possible discharging path (accounting for driver correlation). For each simulation the current waveform for every wire segment in the net is obtained. Based on these current waveforms the peak, average and RMS current densities are calculated for each wire segment and verified against the current density limits. The simulation step is more time consuming than the filtering step, but much more accurate.

Once analysis has been performed on all nets feedback to a designer is provided in several forms: textual reports, GDS II overlays, and a graphical form. As in power grid analysis, graphical feedback is provided in such a way that either current densities or the MTTF for each net is color-coded which allows for an easy identification for critical structures in the design. The overall MTTF for the design is calculated as well.

## 3. Case studies

To demonstrate the application of full-chip reliability analysis techniques presented here a design is selected to study their application. The design selected is a 1.7 million-transistor multimedia chip. The design was modified to illustrate a variety of failure mechanisms. In addition, the chip is analyzed using an operation frequency of 200 MHz. Figure 1 shows an identification of the 10 major blocks in the design. The power grid contains approximately 4 million resistors. No reduction was performed in extraction and contacts are not clustered.

Both static and dynamic power grid analyses are performed on this chip. Power grid analysis is based on analysis of a grid using average currents and the primary analysis of interest is electromigration analysis. Once the transistor-level representation of the design is simulated to determine the average transistor tap currents, the power grid is analyzed. Figure 2 shows the result of applying electromigration analysis to the chip and generating a graphical report. This figure shows the VDD grid for the chip in a variety of levels of gray. Different gray levels indicate different metal routing layers (graphical reports are normally colored, but are gray here due to limitations in printing). Errors are superimposed on the grid as bright white spots in this case. A variety of errors are flagged. Two sets will be discussed.

The first set of errors is along the upper portion of the chip where block B2 is connected to the peripheral power ring. The second set of errors occurs in block B5 on the right side of the chip. The source of these errors demonstrates why reliability analysis of power grids must be performed at the full chip level.

These two sets of errors are due to the currents resulting from the assembly of the chip blocks. Block B3 is not connected to the power rail directly, but is supplied by power through blocks B2 and B5. Since the two blocks are not designed to route power to block B3, their power grids conduct much more current than expected, yielding electromigration violations. The violations at the top of block B2 are in the vias connected to metal 3. Additional current flows through these vias into the vertical metal 3 lines and eventually diverts horizontally to the right to block B3. Although a vertical power bus is provided between blocks B2 and B3, the larger

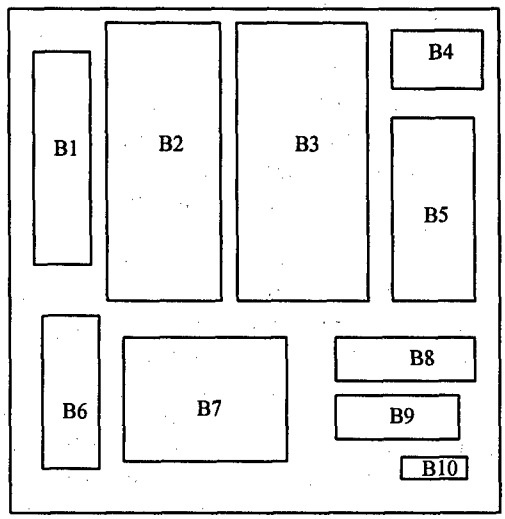



Figure 1. Blocks of the design in the case study.

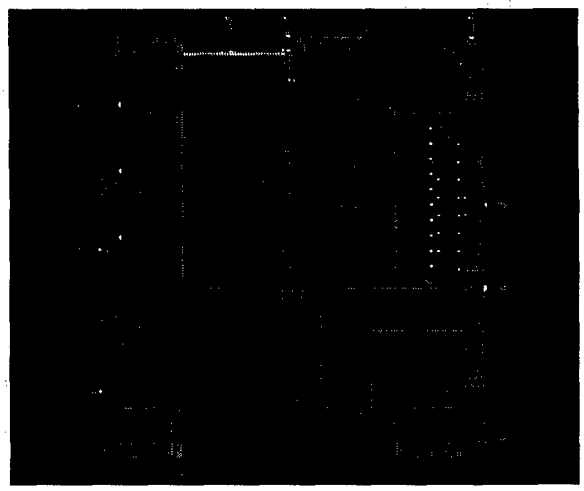


Figure 2. Initial electromigration violations in chip.

effective bus width over block B2 is the primary conduit of current.

The errors in block B5 are also at vias due to excessive current routed through the block. There is a power rail routing power around block B5 to block B3, but due to the power routing in block B5, the path through block B5 is less resistive than the path around the block. This is a case of block power routing drastically altering expected full chip currents.

As an experiment in improving the power routing of the chip, additional wires are added to the chip to connect the top of the design to the bottom. This will allow some current for block B3 to be supplied from the bottom of the chip rather than the top. This should reduce the electromigration violations. The

chip is reanalyzed and the results are shown in Figure 3.

Something quite unexpected has occurred in this experiment; new electromigration errors have been introduced in blocks B7 and B9. The addition of power routing between the top and the bottom of the chip has drastically altered power currents. As expected current is supplied from the bottom. However, the additional current follows paths through blocks and introduces new violations in the design far away from the modifications, rather than only reducing the previous errors. Designers experienced in power routing would not have predicted this behavior. This illustrates why designers need reliability analysis.

Another typical electromigration error is shown in detail in Figure 4. This figure illustrates a 2 by 5 via cluster between metal 3 and metal 2 in a design. Current enters the cluster on metal 3 from the right, passes through the cluster, and exits on metal 2 to the bottom. Shading indicates the relative magnitude of current through each via. In this case the power grid analysis shows excessive via currents crowding to the lower and right corner of the cluster. If contacts were clustered these violations might have been missed.

Companies designing high-speed chips, particularly microprocessor designers, have recognized the value of this power grid reliability analysis. However, this analysis has also been applied to a variety of ASIC chips and identified reliability problems.

In addition to power grid analysis, signal reliability analysis is applied to the design of Figure 1. We selected the clock signal distribution tree (network) to illustrate the analysis. The clock tree is composed of 6700 nets and spans the chip. The source of the clock is a driver at the bottom of the chip and 6699 nets and signal buffers are distributed throughout the chip. Signal analysis takes place in two steps. The first step performs the filtering as described above. The results of this filtering step are shown in Figure 5. As in the power grid figures, metal lines are shown in various gray scales. Metal segments that failed the filtering are shown in white. As expected, the largest clock nets have been flagged because of both their high current and high capacitance due to both loading and parasitics. In addition, a variety of other nets throughout the chip have been flagged. In total, 90 nets of the original 6700 have been identified as potential reliability problems (which we call reliability critical nets) and passed on to the simulation step of analysis.

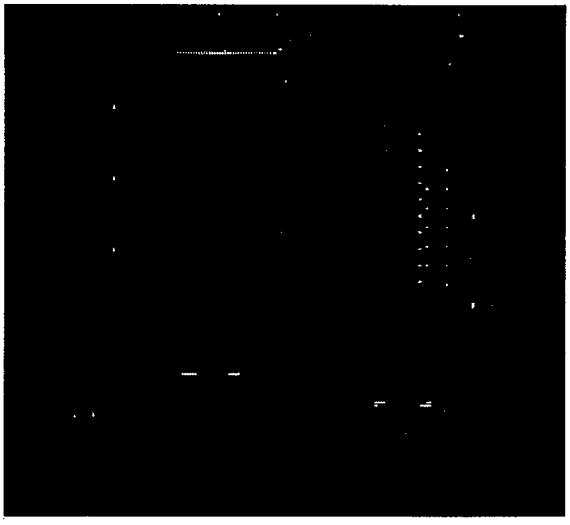


Figure 3. New electromigration violations in chip.

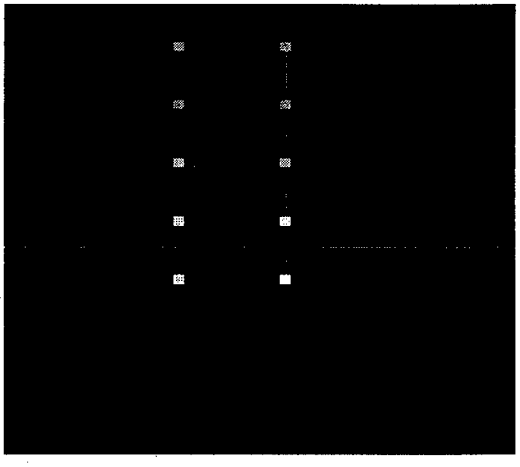


Figure 4. Corner crowding electromigration error.

In the simulation step the 90 potential reliability problem nets are simulated to determine the actual average, peak, and RMS current densities for each wire segment on each net. These current densities are then compared to specified limits and an overall MTTF is reported. The results of the simulation step are shown in Figure 6 for RMS analysis. In this case 6 of the 90 nets are identified as being close to RMS current limits when this chip is operated at 200 MHz. Examination of these errors proves informative.

One error occurs in the lower right side of block B2. A clock driver driving 6 additional large clock drivers nearby caused this error, but several jumper wires in the routing are too narrow for the high currents through the wires.

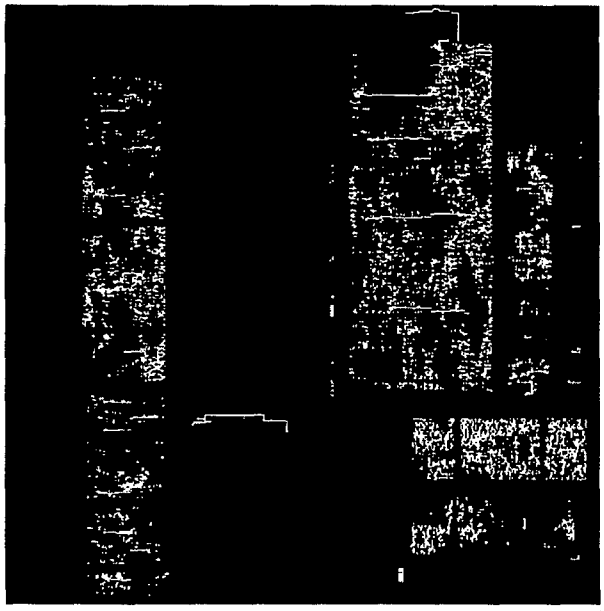


Figure 5. Signal electromigration filtering result.



Figure 6. Signal electromigration simulation result.

A set of errors occurs in block B8 of the design. These violations also occur do to a large driver and larger loading. These errors seem to occur because the clock routing in block B8 is designed differently than the rest of the chip in that they have a higher loading per driver than other blocks. The signal line reliability analysis has therefore succeeded in identifying the signal lines likely to fail in reliability.

## 4. Conclusions

In summary, this paper presents a series of techniques, which provide an engineering solution to performing full-chip reliability analysis for both electromigration and Joule heating on both power grids and signal nets. These analyses make it possible for designers to determine where the reliability weak spots are in their design after all parts have been assembled. The location of failures cannot be determined until the chip is considered as a whole because adjacent blocks in a design have a significant impact on each other in terms of power flow. In addition, signal routing requires novel filtering mechanisms and simulation capabilities in order to give designers the confidence they need before fabricating a design. These analyses will be required in future designs because technology trends increase the probability of reliability failures in chip designs.

## 5. Future design optimization

This type of reliability analysis not only lends itself to producing unprecedented levels of performance, yet maintaining reliability, there are a series of "experiments" that can be performed to obtain answers to longstanding questions. For instance, the question of failure criterion is often posed and usually not answered adequately. At one time, an open circuit was required to define failure, but with the advent of refractory "shunt" layers this definition has become inappropriate.

Clearly, a resistance increase is more appropriate than an open. Recently, it has become common practice to use a 10% resistance increase as a failure criterion. However, clearly, there are circuits that can withstand a 10% resistance increase with no adverse effects and others where an increase of the resistance of this size can produce an unacceptable speed shift.

Using this technique, resistance changes as a function of time can be programmed into the model as a function of the current that is flowing through a specific circuit element. This can then be related to the performance changes related to the way the current is flowing. From "experiments' such as this on real circuits, the answer of what should be used as a failure criterion and what the test structure should look like can be answered to a degree heretofore impossible. Such experiments are planned in the near future.

## 6. References


[1] B.-K. Liew, P. Feng, N.W. Cheung, and C. Hu, "Circuit reliability simulator for interconnect, via, and contact electromigration," *IEEE Tr. Electron Devices*, vol. 39, no. 11, pp. 2472, 1992.

[2] J.R. Lloyd and J. Kitchin, "The electromigration failure distribution: the fine-line case," *J. Applied Physics*, vol. 69, pp. 2117, 1991.

[3] W. Feller, *Introduction to Probability Theory and its Applications*, vol. I, New York: John Wiley and Sons, pp. 166, 1957.

[4] J.R. Black, "Mass transport of aluminum by momentum exchange with conducting electrons," in *Proc. 6th Annual IRPS*, pp. 148, 1967.

[5] M. Shatzkes and J.R. Lloyd, "A model for conductor failure considering diffusion concurrently with electromigration resulting in a current exponent of 2," *J. Appl. Phys.*, vol. 59, pp. 3890, 1986.

[6] M.A. Korhonen, P. Borgesen, K.N. Tu, and C.-Y. Li, "Stress evolution due to electromigration in confined metal lines," *J. Appl. Phys.*, vol. 73, pp. 3790, 1993.

[7] A.S. Oates, E.P. Martin, D. Alugbin, and F. Nkanash, "Comparison of Al electromigration in conventional Al alloy and W-plug contacts to silicon," *Appl. Phys. Lett.*, vol. 62, pp. 3273, 1993.

[8] J.M. Towner and E. van de Ven, "Aluminum Electromigration Under Pulsed DC Conditions," in *Proc. 21st Annual IRPS*, pp. 36, 1983.

[9] *Thunder Manual*, Simplex Solutions, 1998.

[10] *Lightning Manual*, Simplex Solutions, 1998.

[11] I.A. Blech, "Electromigration in Thin Aluminum Films on Titanium Nitride," *J. Appl. Phys.*, vol. 47, pp. 1203, 1976.

[12] *Hail Manual*, Simplex Solutions, 1998.

PERGAMON

Microelectronics Reliability 38 (1998) 861–868

MICROELECTRONICS RELIABILITY

# An analysis of the quality and reliability supplement to the SIA Roadmap

DR. Robert W. Thomas

Technology Experts Network, 104 Cedar Street, Rome, N.Y. 13440

**Abstract**

This is an analysis and summary of the Quality and Reliability Supplement to the 1997 U. S. Semiconductor Industry Roadmap. The key reliability issue with the semiconductor industry over the past 5 years has been the lack of improvement in the Early Life (EL) failure rates. Reasons for this stagnation are discussed and the constraints to quality and reliability improvement in general are presented from a survey conducted by the Q&R team. Major portions of this paper are extracted from the Feb. 1997 Q&R National Technology Roadmap
© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

The Roadmap is the product of all sectors of the U.S. semiconductor technology base. This base includes industry, universities, and government organizations such as National Institute of Standards and Technology (NIST); the National Science Foundation; and the Departments of Commerce, Defense, and Energy including federal and national laboratories. The industry consortia, Semiconductor Research Corporation (SRC), and SEMATECH, are key participants. Broad acceptance of the Roadmap has rapidly materialized, leading the SIA's (Semiconductor Industry Association) Coordinating Group (RCG) to adopt a proposal that the Roadmap be referred to as the National Technology Roadmap for SC's.

The Road map creation process enables a common vision among industry, academia, and government. This vision must be nurtured to allow progression from academic research through manufacturing and commercial products. The Roadmap also provides a framework for guiding R & D; all relevant segments of the national R & D base can be efficiently enlisted to meet the increasingly complex technology needs of the semiconductor industry. (See Table 1, Reliability Related Technology Constraints Prioritization) It demands a growing urgency to effectively fund R & D in search of innovative solutions; it is designed to build a culture of "urgency without crisis".

The updated 1997 National Technology Roadmap for Semiconductors is available for viewing on the Web. The addresses are:
http://notes.sematech.org/1997pub.htm
http://notes.sematech.org/public/97pelec.htm

The Sematech Quality Council sponsored the Reliability Technology Advisory Board that prepared the 1997 "Quality and Reliability National Technology Roadmap Supplement. The supplement which will be analyzed in this paper will be available for viewing on the Web shortly.

## 2. Stagnation of Early Life Reliability.

A key issue carried forward from the 1994 Quality and Reliability Roadmap is the shortfall in defect density learning for yield compared to a better improvement expected by customers in quality and reliability from 1992-1998. There is some divergence in outlook of what it will take to satisfy customers in the year 2000 for quality and reliability. See Table 2 " Y/E 2000 Base Customers Expectation Table" and Table 3 "Quality/Reliability Learning Table". Some believe customer expectations for quality and reliability can be achieved by maintaining the current levels in test escapes (SPQL), early life (EL), and end of life (EOL) failures at Y/E 2000. Others project continued reduction to meet the customer challenge of 1.0 -0.1 FITs in the strategic time frame. What is evident is that the projections made in 1994 for 1997 have been achieved in Test Escapes, and End of Life, but not for Early Life reliability. Indeed, for EL, the learning curve has flattened out with no improvement between 93-97.

Whether continued learning to 1.0 - 0.1 FITs, or no further learning to meet customer satisfaction in Quality and Reliability is accepted as the objective, either expectation will be difficult to achieve with the continued stretching of existing processes and

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00153-X

the rapid introduction of the most significant change in material set that the industry has seen (Cu, LoK dielectrics, Oxi-nitride gates, etc.).

There is a direct connection between defect density and wafer yields. A direct connection between field failure rates and defect density has been established by Kuper and van der Pol at Philips in their 1996 International Reliability Physics Paper. (1) In the range of 80% yield, Philips was achieving single digit Early Failure ppm rates. This reliability value is neither good or bad. It is more an economic sweet spot for the technology involved rather than a concerted effort to achieve 1 ppm Early Life reliability. Customers need to design systems with this fact in mind.

The economic competitive payoff is the main driver for leading edge technology, but with the very important impact of reduced Early Life Failures. With yields of greater than 98% some semiconductor manufacturers have been able to demonstrate near zero Early Life failures, and near the 1998 End of Life goal of 0.1 FITs. This achievement was not with out a paying a high price. It is estimated that the cost of getting to 98% yield was greater than 1 billion dollars! Very few companies have the capital resources to undertake such a program. This explains why the defect reduction goals have not kept up with customer expectations. Only through volume sales of state-of-the-art microprocessors has this been economically possible. It is interesting to note that the defect reduction effort was driven by an economic incentive and that the by-product was a Early Life reliability improvement. It proves that managers are motivated to spend large sums of money on yield improvements but not to achieve high reliability. This is very important for customers to understand. Asking a manufacturer for 0.1 FIT reliability is a futile request. They simply can not commit the enormous resources necessary to meet a customers aggressive reliability requirement. It is competitive economic forces that drive the semiconductor industry to improve yields. It also explains why there is so much variation from technology to technology and manufacturer to manufacturer in Early Life reliability.

## 3. Major Reliability Constraints

Discussed below are the major Reliability Constraints that were listed in the Q &R Roadmap. The major and minor constraints are listed in Table 1. Each of the major constraints will be discussed the in next nine subsections. The major constraints are listed in an averaged priority order. Note that in Table 1 each company has its own priority listed. Very few are in agreement and for good reasons. The technology mix being produced by each company is different. The priorities for memories are much different than for microprocessors. This also explains why it is very difficult to be a leader in more than one technology area. One single manufacturer simply cannot address all the constraints at one time. It should also be obvious that a small company making linear technology does not have the same capital resources or requirements that a major microprocessor manufacturer would have.

Gate dielectrics and electromigration are listed as the two major reliability issues. In these cases, the concern is maintaining the wear out or End of Life reliability goals of 15 - 25 years. Although in some applications, the technology applications are moving so fast that lap top reliability goals of 3 - 5 years are practical. Because of the uncertainty in our prediction techniques most manufacturers have a minimum End of Life goal of 10 years or more. Gate dielectrics are also very sensitive to defect density. The dielectrics are becoming so thin that single atom contamination can reduce the gate oxide dielectric strength.

Table 1 is a manufacturer's response to a survey conducted by the Quality Council, Reliability Technology Advisory Board projecting the Reliability Related Technology Constraints for the year 2000.

It is interesting to see how differently the various manufacturers view their Reliability Constraints. (e.g., what is the biggest reliability challenge for them) These priorities are a function of how much effort and money these companies have spent on reducing their #1 challenge and the product which they make. A specialist memory company will likely have different priorities than a microprocessor manufacturer.

The last column on the right is a numerical sum of the companies ranking for each category.

### *3.1 Gate Dielectric Breakdown*

#### *3.1.1 Problem Statement:*

The susceptibility to gate dielectric breakdown will increase due to: thinning of the dielectric; small increases (after correcting for band bending) in the gate electric field; increase in the gate area/die (especially if large decoupling capacitors are needed); and increased sensitivity to post-gate processing. Furthermore, as the gate oxide thickness is scaled below ~50 $A^{o}$ the gate conduction mechanism will change from Fowler-Nordhiem tunneling, which we now understand pretty well, to direct tunneling. This could impact, positively or negatively, the reliability of the gate dielectrics.

**Table 1. Reliability related technology constraints (Year 2000 projection) prioritization**

**SILICON**

| | a | b | c | d | e | f | g | h | i | TOTAL |
|---|---|---|---|---|---|---|---|---|---|---|
| Gate Dielectric Reliability | 1 | 4 | 1 | 3 | 5 | 1 | 1 | 5 | 4 | 25 |
| Electromigration | 5 | 5 | 3 | 6 | 6 | 5 | 3 | 13 | 1 | 47 |
| ESD | 6 | 6 | 8 | 2 | 8 | 3 | 5 | 9 | 2 | 49 |
| Multi Level Metal/Dielectric Integrity | 12 | 8 | 10 | 5 | 1 | 10 | 2 | 3 | 8 | 59 |
| Hot Carriers | 9 | 1 | 2 | 8 | 12 | 6 | 4 | 11 | 7 | 60 |
| Defectivity, Cleanliness | 4 | 9 | 4 | 1 | 15 | 7 | 13 | 1 | 12 | 66 |
| Wafer Charging/Antenna Effects/Ultra Thin Oxides | 2 | 17 | 15 | 16 | 4 | 2 | 7 | 2 | 3 | 68 |
| Noise Margin/Coupling | 13 | 2 | 5 | 4 | 7 | 9 | 14 | 12 | 6 | 72 |
| Latchup | 15 | 7 | 9 | 10 | 13 | 4 | 6 | 4 | 11 | 79 |
| Leakage Isolation | 8 | 14 | 6 | 11 | 10 | 11 | 9 | 10 | 15 | 94 |
| Tools for Reliability Checking | 3 | 12 | 17 | 17 | 9 | 8 | 12 | 7 | 9 | 94 |
| Package Induced Failures | 14 | 13 | 14 | 15 | 3 | 15 | 10 | 8 | 10 | 102 |
| Soft Error (Single Error Upset) | 11 | 16 | 16 | 13 | 2 | 14 | 16 | 14 | 5 | 107 |
| Cost Effective Reliability & Qualification | 7 | 11 | 7 | 114 | 14 | 17 | 8 | 17 | 13 | 108 |
| Mixed Signal Requirements for Transistor Matching | 17 | 15 | 13 | 2 | 16 | 12 | 11 | 6 | 17 | 109 |
| Performance/Power/Reliability Trade Offs | 10 | 10 | 11 | 7 | 11 | 16 | 115 | 15 | 16 | 111 |
| Increasing Tj | 16 | 3 | 12 | 9 | 17 | 13 | 17 | 16 | 14 | 117 |

1 = Highest Priority

*3.1.2 Key Issues:*

Measuring TDDB requires large sample sizes and may not be measuring the failure mechanism that occurs during normal device operation.

There is still uncertainty about the model for field acceleration of dielectric breakdown under DC biases.

The transition to direct tunneling could introduce new failure modes.

Gate dielectric reliability will be increasingly sensitive to the complete process flow.

There are other issues, such as diffusion through the dielectric from p channel poly gates and band bending.

*3.2 Metal Interconnect*

*3.2.1 Problem Statement:*

Present and planned reductions of metal-interconnect dimensions are reducing and making more uncertain the margin for error in designing and manufacturing metal interconnects of adequate reliability. Significant efforts are needed to optimize and measure reliability performance. Remember, the sigma of our grain boundary models becomes bigger as we reduce our line width.

*3.2.2 Key Issues:*

Gaps and deficiencies exist in the tool set for measuring reliability of interconnects. They exist, for example, in evaluating vias, in the effect of pulsed-current stresses, and in characterizing the susceptibility to stress voiding. Gaps also exist in characterizing the early reliability of the interconnect system. To help fill these gaps requires a better understanding of the effect of accelerated stress conditions on the metal system so that more accurate estimates can be made of the actual acceleration of the stress test and an assurance that the conditions of the stress test are relatable to use conditions.

Copper-based metallizations offer the advantages of lower resistivity and higher potential resistance to electromigration if processing problems can be overcome. Work is needed first to solve problems related to copper deposition, containment, and integration with adjacent materials.

3.3 ESD

*3.3.1 Problem Statement:*

The continued Human Body Model (HBM) requirement on ESD robustness of >2KV, the Machine Model requirement of >200V, and the Charged Device Model (CDM) of >1KV will put constraints on the ability to scale to smaller geometry's and to operating devices at lower voltages. New protection circuit designs will have to account for process technology variations and

design related cases to different fabrication facilities. Additionally, with increasing pad counts and higher density circuits, space available for effective ESD will be at a premium.

*3.3.2 Key Issues:*

Silicon technology has a strong influence on ESD capability. Designs which function well in one technology do not necessarily show similar performance in newer sub-micron processes. Circuit elements need to withstand heating effects, sink large amounts of current, and not be damage by high electric fields. Key process parameters which influence ESD performance need to be identified and characterized. Unfortunately, the implementation to ESD protection is iterative. Capability of the process is characterized on test structures which provide design guides for protection circuitry. This approach which relies on empirical results as opposed to using accurate simulation tools does not allow for robust technology development. Also limiting ESD issues has been the lack of agreed to standards and test methods. Differences in tester capabilities and their calibration can result in different ESD threshold voltage values. Test methods also influence results based upon what combination of pins are selected for ESD testing.

*3.4 Multi Level Metal/Dielectric Integrity*

*3.4.1 Problem Statement:*

Aggressive scaling and increased performance requirements of semiconductor devices is resulting in: an increased number of metal interconnect layers will impose additional constraints to the selection of materials and their integration, Smaller Via/interconnect line features sizes, often involving many different materials and interfaces within a 1 $um^3$, will aggravate interface problems, new materials like Low K dielectrics and Low Resistance metals to achieve the projected performance targets are untested, and increased die sizes and novel assembly options like C4 (Ball Grid Array) may result in increased mechanical and thermal stresses.

*3.4.2 Key Issues:*

The impact of a high mechanical stress environments into the performance and reliability of silicon semiconductor devices. This will require improved experimental and modeling techniques for materials characterization (Thin film mechanical properties and interface adhesion strengths) in the micron/submicron regime, including in and around vias. This issue is real in that current product that is being shipped with voids under the vias. They have been characterized and shown to not be a reliability risk but they really look bad.

The reliability of new dielectrics and low resistance metals as interconnects. 8 levels by 2010.

*3.5 Hot Carriers*

*3.5.1 Problem Statement:*

Aggressive performance requirement of MOS devices will require continual channel length reduction and improved Source-Drain engineering. Even with the simultaneous reduction in power supply voltages as the devices scale down, the Hot Carrier effect is most pronounced with the reduction of channel length and the thinning of gate dielectric. Hot Carrier effect can impact both device design and technology development as it lowers device current drive and thus limits device speed performance. Further, scaled technology will be sensitive to small drift in parameters.

*3.5.2 Key Issues:*

Accurate Hot Carrier degradation modeling and simulations are critical since over-estimation of this effect can put unnecessary constraints on device design and prevent it from realizing its maximum speed potential, whereas under-estimation of this effect can jeopardize device long-term reliability. An accurate model to correlate wafer level Hot Carrier reliability testing with actual operating conditions in order to optimize device performance without paying the penalty for lesser device reliability is essential.

The conventional $I_{sub}$ (max) measurement techniques are not sufficient. A unified industry-wide Built-In reliability approach is needed. If the first time we see an impact of Hot Carrier Injection induced degradation is during process development and product qualification we are too late. We need this information built into our design libraries. Additionally, we need an industry-wide accepted technique.

*3.6 Wafer Charging/Antenna Effects/Ultra Thin Oxides*

*3.6.1 Problem Statement:*

Gate oxide and transistor scaling will require a better understanding and a reduction of the effects of process-induced charging damage. High power and high density plasmas, with the effects of such schemes as magnetic confinement, will only serve to enhance the charging issue. Common processing steps such as plasma enhanced

deposition, etching, ashing, and implantation may become significant contributors to wafer charging.

*3.6.2 Key Issues:*

Processing Issue: The thin oxides are expected to be less tolerant to wafer charging. For example, if the poly (or contacting metal) charges to the point where the field in the gate oxide reaches 10 Mv/cm, severe damage can occur to the gate oxide.

Design Issue: The widely used design scheme for antenna protection (diode protection) is rapidly becoming ineffective. When the antenna ratio (area of thick-oxide / area of thin oxide) exceeds a critical value (~ 100/1), the designer must tie the gate to a diode for protection. The diode provides full protection (independent of polarity of charging) only if the diode in parallel with the gate oxide can breakdown before the 10 Mv/cm across the oxide is reached. For diodes that breakdown at 10 V, the diode can provide protection only for oxides >100 $A^0$. Thus, for certain process steps, standard diode protection may not be sufficient protection for <100 $A^0$ oxides.

*3.7 Increase in Max. Junction Temperature*

*3.7.1 Problem Statement:*

The current 100 MHz devices would be expected to generate about 9 Watts at 3.3 V operating voltage. For a 150 MHz devices one would expect about 12.5 Watts at 3.3 V. If we lower the operating voltage to about 2.5 V and we have effectively lowered our power generated to about 10. Watts. Yet its higher than the previous generation (the 100 MHz at 3.3 V) by 1 Watt. So we have a faster part at a lower voltage but generate more power even though the voltage was scaled down. With future technologies (projecting to year 2000) the number of transistors will scale up and frequencies will scale up too. We will scale down voltage only. Will we be able to scale voltage down fast enough to keep a control on power consumption? No, not according to the simple example above which did not even factor in increases in the number of transistors.

*3.7.2 Key Issues:*

Power generated scales up faster with frequency than we can/"will" scale down with operating voltage.

Increase in power means just one thing ... increase in $T_j$ max. Device characteristics would definitely be impacted, diode breakdown voltages would be impacted, worst case - maybe even latch sensitivity. Designers would have to design the part over a wider temperature range.

Need to study the effect of increasing layers of metallization on $T_j$ max. and, the effect of self-induced heating of upper layers of metallization on electromigration.

*3.8 Soft Error (Single Error Upset)*

*3.8.1 Problem Statement:*

The downward scaling of the power supply is expected to continue to be driven by improvements in the Performance/Power ratio. The susceptibility of memory circuits and internal microprocessor caches to single event upsets/soft errors is expected to increase dramatically with decreasing $V_{CC}$.

*3.8.2 Key Issues:*

Memory Issues: Memory devices depend on sufficient charge being stored in the cell during the writing operation that it can be reliably be detected during the reading (sensing) operation. It is well known that ionization-producing charged-particles such as alpha particles (emitted from trace amounts of uranium and thorium found in the packaging materials) can induce charge loss in the cell (data state upset). More recently, it has been shown that cosmic rays (the high energy neutron component) can have similar effect on memories. The later mechanism is difficult to prevent and may enhance the need for error correction.

Logic Issues: Although error detection/correction is common in memory systems, it may impose significant performance and/or density penalties for caches in microprocessors. Furthermore, aggressive scaling of $V_{CC}$ along with the dramatic reductions in capacitance needed for very high speed logic operation raise the question of the single event upset/soft error susceptibility of logic circuits.

*3.9 Defectivity, Cleanliness*

*3.9.1 Problem Statement:*

Clean processes are key to developing and maintaining the process under control. They are critical to ensuring that wear out phenomena do not significantly occur during the useful life of the component and that defect densities are at acceptable levels. The projections for increased die size, and reduced reliability fallout expectations by customers, all require an accelerated reduction in defect density.

*3.9.2 Key Issue:*

Defect reduction/yield improvement is a very expensive and time consuming process. The semiconductor equipment industry will play a major role in reducing defects during production

**Table 2**
**Base Customer Expectations at 85C**

| Market Segment | Mfr. -- Year -- | A 97-00 | B 97-00 | C 97-00 | D 97-00 | E 97-00 | F 97-00 | G 97-00 | H 97-00 | I 97-00 |
|---|---|---|---|---|---|---|---|---|---|---|
| Memory | | | | | | | | | | |
| | SPQL (dpm) | | | 70-30 | 5-25 | 50* | | 20 | 3.4-25 | 50-3 |
| | EL (dpm) | | | 100-40 | 100-350 | 200* | | 50-25 | 350 | 600-3 |
| | EOL (FITS) | | | 30-15 | 3-10 | 150* | | 3-260 | 25 | 50-3 |
| ASICS | | | | | | | | | | |
| | SPQL (dpm) | <100 | | 70-30 | <10 | 50 | 10-100 | | 50-500 | 10 |
| | EL (dpm) | <5000 **** | | 100-40 | 175-700 | 200 | 50-1000 | | 54-100 | 5 |
| | EOL (FITS) | <100 | | 30-15 | 5-20 | 150 | 10-100 | | 8-160 | 1 |
| Analog | | | | | | | | | | |
| | SPQL (dpm) | | <5-<3*** | | | | | | 3.4 | |
| | EL (dpm) | | <220-<100 | | | | | | 1-15 | |
| | EOL (FITS) | | <5-<3 | | | | | | <10 | |
| Hi End uP | | | | | | | | | | |
| | SPQL (dpm) | <100 | | 140-60 | <10 | 50 | | 10-100 | 50-500 | 5 |
| | EL (dpm) | <1000 **** | | 200-80** | 175-700 | 200 | | 100-1000 | 54-100 | 5 |
| | EOL (FITS) | <100 | | 60-30 | 5-10 | 150 | | 10-300 | 8-160 | 5 |
| Low End uP | | | | | | | | | | |
| | SPQL (dpm) | | | 70-30 | <10 | 50 | | 15-200 | 50-500 | 5 |
| | EL (dpm) | | | 100-40** | 700-2800 | 200 | | 100-2000 | 54-100 | 10 |
| | EOL (FITS) | | | 30-15 | 20-80 | 150 | | 125-500 | 8-160 | 3 |
| Auto | | | | | | | | | | |
| | SPQL (dpm) | | | | | | | 0 | | 100-5 |
| | EL (dpm) | | | | | | | 100 | | 200-5 |
| | EOL (FITS) | | | | | | | N/A | | 20-5 |
| Communications | | | | | | | | | | |
| | SPQL (dpm) | | | | | | | 55 | | 5 |
| | EL (dpm) | | | | | | | 1000 | | 5 |
| | EOL (FITS) | | | | | | | 100 | | 10 |

Key: SPQL = Shipped Product Quality Level (Test Escapes), EL = Early Life Failures, EOL = End of Life, (dpm) = defects per million, (FITS) = Failures In Time
Notes: EL= 1 yr at 85C, EOL = 10yrs at 85C. * At 70C not 85C for SPQL, EL, & EOL, ** In FITS not dpm, ***At 55C not 85C, **** W/O Burn-in,***** For ASICS and DSP's, Lifetime is 25 years for telecom, 15 years for remote terminal.

**Table 3**
**Quality and Reliability Learning/Customer Expectancy**

| Quality and Reliability | Actual 1993 | Projected 1997 | | Actual 1997 | Projected 2000 |
|---|---|---|---|---|---|
| SPQL (dpm) | | Memory | 3.4-25 | 3.4-70 | 3.4-50 |
| | | ASIC | 10-100 | <10-500 | 10-500 |
| | 25-2000 | High End u | 5-500 | <10-140 | <10-500 |
| | | Low End u | 5-50 | <10-70 | <10-500 |
| | | Auto | 5-300 | 0-200 | 0-10 |
| | | Comm | 5-100 | 10-25 | 5-100 |
| EL (dpm) | | Memory | 3-350 | 50-350 | 3-500 |
| | | ASIC | 5-1000 | 50-700 | 5-1000 |
| | 50-3000 | High End u | 5-1000 | 54-1000 | 5-1000 |
| | | Low End u | 10-2800 | 54-2800 | 10-2800 |
| | | Auto | 5-1000 | 100-500 | 5-100 |
| | | Comm | 5-500 | 50-1000 | 5-1000 |
| EOL (FITS) | | Memory | 3-25 | 3-260 | 3-260 |
| | | ASIC | 1-150 | 5-150 | 1-150 |
| | 20-275 | High End u | 5-300 | 5-150 | 5-300 |
| | | Low End u | 3-120 | 8-150 | 3-500 |
| | | Auto | 5-100 | 20-700 | 5-50 |
| | | Comm | 10-600 | 10-100 | 10-100 |

Note: The low end of the Early Life Failure range has remained constant from '93 to '97 while Test Escapes and End of Life have shown 4X improvement and that these attributes of reliability vary greatly with technology. One really must be very careful in how device reliability is treated in systems development.

(3). Indeed they have assumed this responsibility, but with a dramatic increase in cost of equipment.

**4. Summary:**

As a smart customer you need to know what the major reliability constraints are in achieving reliable product performance. The costs of system down time, like the major network outage that occurred in April in the U.S. are very expensive. One customer lost $10 million dollars in one day. We don't know what caused the outage but Early Life failures do occur. The industry has not been able to improve the Early Life failures in the last 4 years because of economic constraints. Can we live with this situation?

There is a need to understand the cost of defect reduction programs for each technology and each manufacturer and take this into account in parts procurement and system design.

**References**

(1) F. Kuper, J. van der Pol, E. Ooms, T. Johnson, R.Wijburg, W. Koster, and D. Johnston, "Relation Between Yield and Reliability of Integrated Circuits: Experimental Results and Application to Continuous Early Failure Rate Reduction Programs", IEEE International Reliability Physics Symposium, 1996, p 17 - 21. Note: An update of this paper was given at the 1996 ESREF conference.

(2) W.R. McKee, H.P. McAdams, E.B. Smith, J.W. McPherson, J.W. Janzen, J.C. Ondrusek, A.E. Hyslop, D.E. Russell, R.A. Coy, D.W.Bergman, N.Q. Nguyen, T.J. Aton, L.W. Block, V.C. Huynh "Cosmic Ray Neutron Induced Upsets as a Major Contributor to the Soft Error Rate of Current and Future Generation DRAMS", IEEE Proceedings of the International Reliability Physics Symposium, 1996, p 1-6

(3) A.K. Sinha, F. Moghadam, R. Mosley, M. Chang, R. Ellwanger, "Emerging Role of Semiconductor Process Equipment to Overcome Device Failure Mechanisms", IEEE Proceedings of the International Reliability Physics Symposium, 1997, p 1-6.

(4) J.S. England and R. W. England "The Reliability Challenge: New Materials in the New Millennium, IEEE Proceedings of the International Reliability Physics Symposium, 1998, p 1-8.

PERGAMON

Microelectronics Reliability 38 (1998) 869–876

MICROELECTRONICS RELIABILITY

# The use of the Focused Ion Beam in failure analysis

D. Verkleij

*Philips semiconductors, PMO department, Gerstweg 2 6534 AE Nijmegen, The Netherlands*

**Abstract**

*Since the introduction of the Focused Ion Beam (FIB), many applications have been developed. This article deals with a stand alone FIB in a failure analysis laboratory where it is used for material characterisation and sample preparation. In this paper examples will be given of FIB applications as used for failure analysis and process monitoring of semiconductor devices*

© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction.

Since the commercial availability of Focused Ion Beam (FIB) systems in the 1980's, applications of this technique have assumed enormous proportions. Started off as a tool for circuit modification in the semiconductor industry, it is now also an essential tool in failure analysis. Even in non-semiconductor applications the FIB finds a wide spread usage. This article deals with a stand-alone FIB in a failure analysis laboratory as it is used for material characterisation and sample preparation. Examples discussed in this article come from daily practise in this laboratory.

Basically, a focused ion beam has three functions. It can be used as a scanning ion microscope. While $Ga^+$ ions strike the sample surface, secondary electrons, ions, photons and neutrals are emitted. Using a proper detector, secondary electrons and secondary ions can be used to produce an image.

Focused $Ga^+$ ions used within a FIB have the ability to remove material from the sample surface. The second function therefore is maskless etching, providing cross-sections which can be *in-situ* imaged.

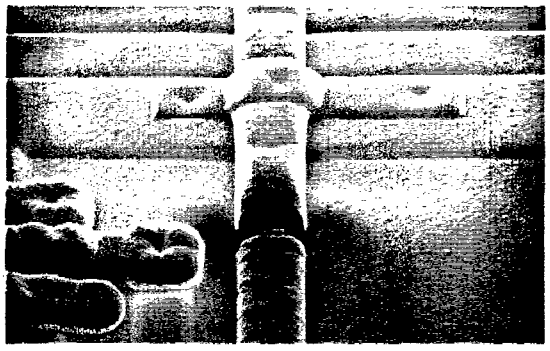


Fig. 1. Device modification.

The third function is material deposition. Using an organometallic compound of - for example - platinum as a source gas, it is possible to deposit a Pt film on top of the sample surface (FIB assisted CVD). By use of an insulator precursor material as a source gas, it is possible to create an interconnect between two metal layers passing through a sandwiched intermediate layer.

Device modification still is an important application of the FIB in the semiconductor industry. Fig. 1 shows an integrated circuit modification with a high resistance cut and a low resistance metal reconnection.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00126-7

Because of scaling down of VLSI structures in the electronic industry it is necessary to improve techniques and methods in order to analyse ever shrinking features. The FIB has turned out to be of great help to meet that challenge. The ability of the FIB to focus the $Ga^+$ beam onto a small diameter, less than 10 nm., provides us with a tool to make site specific cross-sections of very small features and image these cross-sections. However, these images provide us with a lot of information. These cross-sections can be used for further examination by Auger, EDX or SIMS, giving a lot of process information. This property of the FIB system has also enabled the preparation of transmission electron microscopy (TEM) specimens.

## 2. Imaging.

There are many similarities between the processes of secondary electron emission due to energetic ion bombardment and secondary electron emission due to energetic electron bombardment. The contrast mechanism for ion induced electron emission images, as obtained with a FIB, is similar to the contrast mechanism as obtained with a scanning electron microscope. The same broad division can be made: Topographic contrast, Material contrast, Channelling contrast and Voltage contrast. The relative importance of each mechanism in the overall contrast formation will be different for images obtained with a FIB in comparison with images obtained with a SEM. With a FIB secondary electron images as well as secondary ion images can be produced.

### *2.1. Topographic contrast.*

The secondary ion or secondary electron yield depends on the incident beam angle. When an ion beam strikes an irregular sample surface, different sections of the surface are oriented with different angles with respect to the incident beam. This gives rise to a varying secondary particle yield. The escape depth of secondary electrons is about 1 - 50 nm. For secondary ions it is < 1 nm. Together with the short penetration range of an incident 30 keV $Ga^+$ beam (10 - 20 nm.), the information depth of an ion induced secondary electron image and secondary ion image is limited [1]. This makes it very sensitive to fine details in the surface topography as shown in Fig. 2. An ion induced secondary electron (IIE) image of a surface showing a contact.

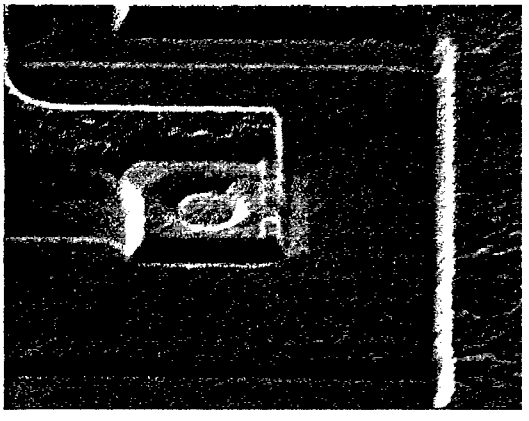


Fig. 2. An ion induced secondary electron image of a surface showing a contact.

### *2.2. Material contrast.*



Fig. 3. On the left, IIE image of a cross-section through a contact. On the right a secondary ion image at the same location.

Secondary particle yield is also influenced by both the elemental composition of the sample and the presence of a contamination layer. Especially secondary ion images are dominated by material contrast due to the large ion yield enhancement or suppression caused by oxide formation and by surrounding materials, a well know effect in SIMS . Ion yield is also affected by beam-induced chemical changes in the sample surface. The combination of ion elemental yield sensitivity, matrix effect and beam induced chemical changes

produce contrast in ion images that are impossible in electron images. See Fig. 3.

*2.3. Channelling contrast.*

An additional effect on the secondary particle yield of polycrystalline material is its dependence on the orientation of grains with respect to the incident beam. When the primary beam is closely aligned with a low index crystallographic direction, primary ions have the ability to channel in the crystal. These ions can travel a considerable distance through the crystal before they are stopped. This reduces the electron yield. Kinetic electron emission depends on the inelastic energy loss and a channelled ion transferring relatively little energy along its path. Electrons generated by a channelling ion are much further from the sample surface, which makes it difficult for these electrons to escape. Such crystals appear dark in the secondary electron images. Others appear bright.

The appearance of crystallographic contrast depends on the removal rate of the contamination layer and the amorphisation rate of the top layer of polycrystalline material. Due to the continuous sputter process, the surface of the polycrystalline material is kept clean. This enhances a good channel contrast. Fig. 4 shows an ion induced secondary electron image of a cross-sectioned aluminium line. For polycrystalline silicon, however, obtaining channeling contrast is much more difficult due to amorphisation of the top-layer.

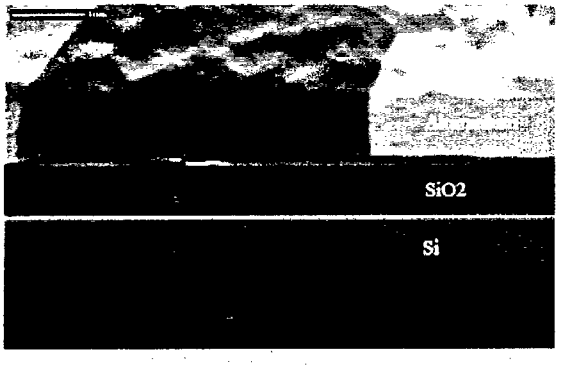



Fig. 4. An ion induced secondary electron image of a cross-sectioned aluminium line. Grains are clearly visible because of channelling contrast.

*2.4 Voltage contrast.*



Fig. 5. On the left an IIE image of a cross-section through a contact. On the right, an image at the same location after implantation of $Ga^+$ ions.

Because of the low energy of the emitted secondary electrons and secondary ions the yield of these particles also depends on local surface potential variations. There exists a huge difference between secondary electron yield and secondary positive ion yield. This results in the tendency of the sample to charge positively, depending on the local conductivity. An effect reinforced by the implantation of primary $Ga^+$ ions. As a result a variation in the contrast of the ion induced secondary electron image shows up. This is a powerful tool in failure analysis because an estimation can be made whether an area is floating or grounded. Due to implantation of $Ga^+$ it is possible to change the local conductivity. After a certain dose of $Ga^+$ ions a silicon nitride layer appears brighter in an ion induced secondary electron image. See Fig. 5.

## 3. Analysis of FIB made cross-sections.

Because of the ever shrinking details of VLSI devices it becomes troublesome to prepare cross-section through located faults by mechanical polishing. Using AES in combination with ion sputtering or EDX in combination with 'reverse engineering' to analyse included particles for example, becomes a very difficult task. Moreover, due to the influence of ion bombardment induced interface roughening in Auger analysis, it is almost impossible to determine in which of the different layers the particles are contained. The elemental composition cannot always be determined unambiguously because of signal interference with the material in which the

particles are included. On the other hand, 'reverse engineering' can lead to contamination which might affect the results or it does not have the required selectivity in material removal.

With modern FIB systems positioning of the $Ga^+$ ion beam on the wafer or device can be done with an accuracy of 0.1 μm. This allows us to make very precise cross-sections of localised faults.

All cross-sections discussed in this article are prepared on an FEI 800 FIB system, using a beam energy of 30 keV. The chamber pressure was better than $10^{-6}$ mbar during milling and $10^{-7}$ mbar when the images were taken. Cross-sections were made in two steps: First milling a hole with a length of 5 - 10 μm. The widths were about 6 μm. and the depths about 4 μm. This was done with a maximum current of 7 nA., using iodine as an enhanced etch gas. Typical milling time was 6 - 8 minutes. Secondly, the surface of the cross-section was cleaned with a current of 70 pA using the line scan pattern. Cleaning takes about 3 - 6 minutes. Imaging the cross-section was done at 3 - 4 pA, 1024 x 1024 pixel resolution and 4 μs. per pixel dwell time. If cross-sections were subjected to further analysis, the surface was again cleaned according to the procedure mentioned above, to get as less gallium implantation as possible.

### *3.1 Auger analysis*

As a first example of the Auger measurement on a FIB made cross-section, the analysis of an included particle is shown. Fig. 6 shows an optical image of a surface distortion caused by an included particle.

The ion induced electron image of the cross-sectioned particle is shown in Fig. 7. Auger measurements were performed on a PHI-670

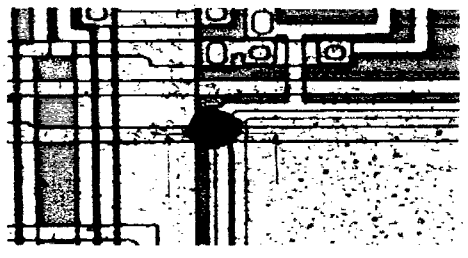


Fig. 6 Optical image of a surface distortion

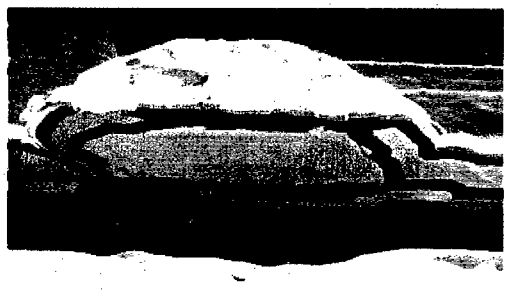


Fig. 7. IIE image of the FIB made cross-section of the included particle shown in Fig. 6.

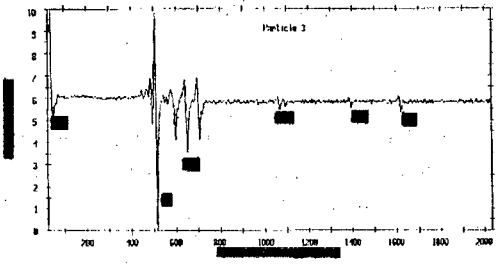


Fig. 8. Auger surface spectrum of the cross-section.

scanning FE Auger nanoprobe. The emission source was operated at 20 keV, 20 nA with a spotsize < 40 nm. Prior to Auger measurements, the cross-section was sputter cleaned for one minute with a focused $Ar^+$ ion beam. Under the chosen conditions the sputter rate was estimated at 100 nm/min. for $SiO_2$. Fig. 8 shows the initial Auger spectrum of the particle. The LMM triplet of Fe shows up clearly. An elemental map of Fe was made to show the distribution of Fe. See Fig.9.

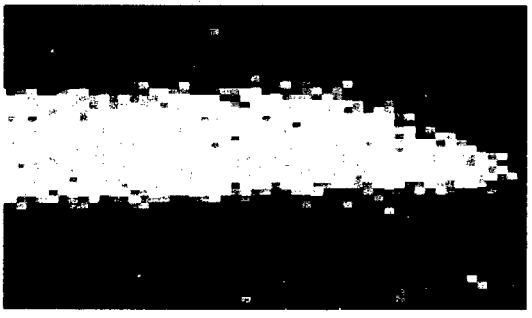


Fig. 9. Fe elemental mapping of the cross-section.

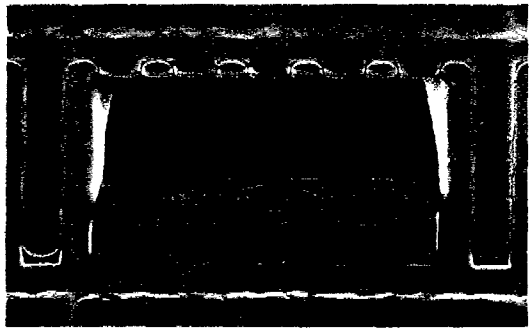


Fig. 10. IIE images of a cross-section through contacts.

A second example is the structural analysis of contacts of a transistor. Fig. 10 shows the ion induced secondary electron image of the cross-section. This cross-section was made under an angle of 45° with respect to the normal of the sample surface. This provides us with an optimum position of the cross-sectioned surface with respect to the Auger detector. Auger measurements were performed on a VG 310 F Auger nanoprobe. The same conditions were used as described above except for the sputter rate. This was estimated at 15 nm/min. for $SiO_2$. To obtain the mapping a dwell-time of 100 ms. was used. See Fig. 11.

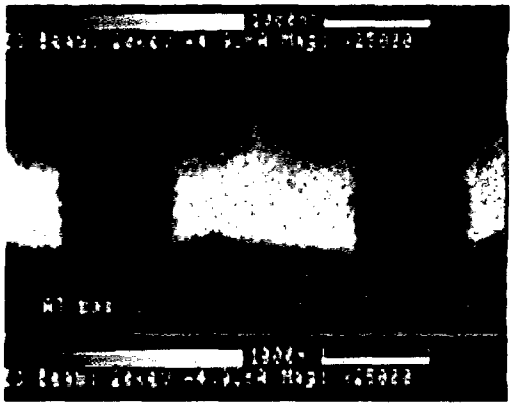


Fig. 11. Aluminium map of one of the contacts.

*3.2 SIMS analysis.*

Nowadays SIMS also is commonly in use in failure analysis. When the material of a localised feature needs to be analysed by SIMS, the following problems may arise.

To analyse a small area on the wafer or device it is necessary to position the SIMS ion beam with an accuracy down to about 0.1 μm. The area/volume to be analysed is very small (0.5 x 0.5 x 0.2 μm.) Such a small analysed volume (number of available atoms) increases the detection limit. These problems could be solved by the use of a TOF-SIMS.

Integration of a quadruple SIMS into a FIB instrument gives access to secondary ion information with high lateral spatial resolution, available when the instrument is in use [2]. The mass resolution and detection limits are limited in comparison with dedicated SIMS instruments. For structural analysis and particle analysis however, it is a very useful tool

As an example, a structural analysis of a short between two metal lines is given. After the cross-section was prepared as described above, chemical mappings were acquired using the SIMS quadruple. A beam energy of 30 keV and a beam current of 70 pA were used. The pixel resolution was 256 x 256 and 2ms. per pixel dwell time. Using these conditions a 60 nm thick layer was removed during SIMS analysis. Fig. 12 shows an ion induced secondary electron image of a FIB made cross-section through a metal 2 plate. The aluminium is locally oxidised to divide the plate into several lines. Because of the voltage contrast the aluminium oxide between the two lines appears dark. The aluminium oxide clearly shows up in the oxygen map, made at the same location, shown in Fig. 13. The oxidised metal 1 line is visible as well.



Fig. 12. IIE images of a cross-sectioned aluminium plate.

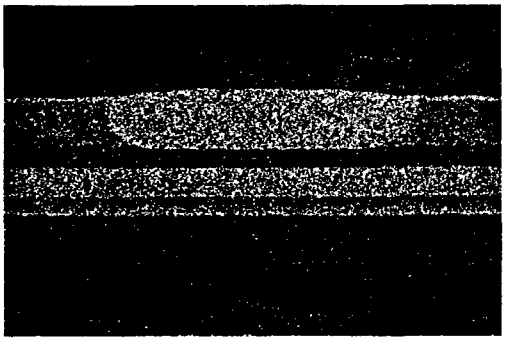


Fig. 13. O map of the cross-section shown in Fig.12.

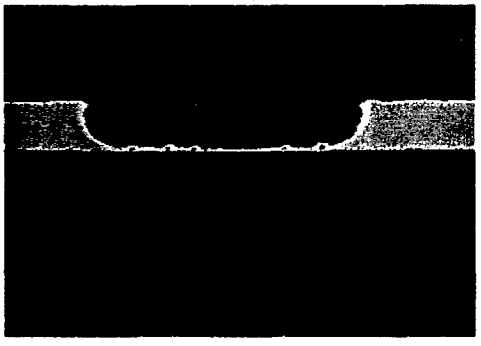


Fig. 14. Al map of the cross-section shown in Fig. 12.

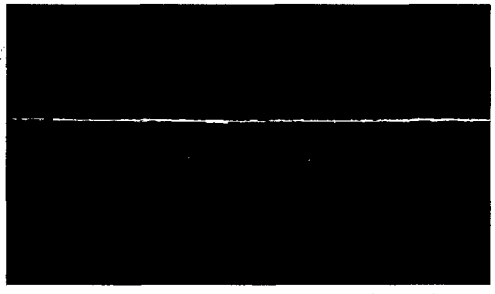


Fig. 15. Hf map of the cross-section shown in Fig. 12.

Fig. 14 shows the elemental map of aluminium. Already visible in Fig. 12 however, this map led to the conclusion that the aluminium was not fully oxidised. Therefore, a short remains between those two lines. Underneath the aluminium, a thin layer of hafnium is deposited. The elemental map of this hafnium layer is shown in Fig. 15.

## 4. Grain size observations.

A focused ion beam is capable of imaging grains in, e.g., aluminium films. Since aluminium grain size distribution affects the performance and reliability of aluminium lines in integrated circuits this capability of the FIB is very useful in failure analysis. Because of its speed to obtain results it is useful as a monitor for Al deposition equipment as well.

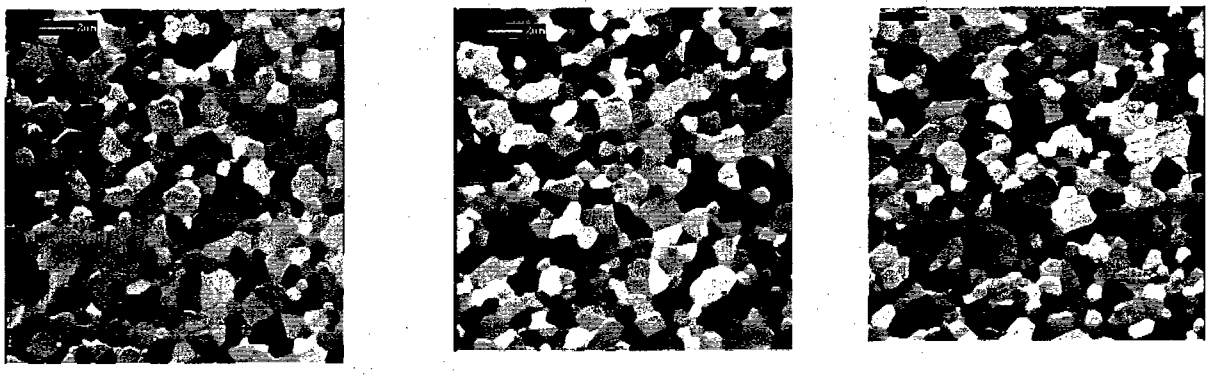
In contrast to a transmission electron microscope, no special sample preparation is necessary. After removing the oxide layer, the grains become clearly visible as a result of channelling contrast. Grain size measurements using a FIB are in good agreement with TEM observations of grain sizes [3]

Fig. 16 shows aluminium grains obtained with a FIB. After removal of the oxide layer (sputtering with 30 keV, 150 pA for 1 min.) grains were observed under three different tilt angles with respect to the primary beam: 15°, 20° and 25°. Pictures were taken using a beam current of 70 pA, 1024 x 1024 pixel resolution and 4 μs per pixel dwell time. The pressure remains at $10^{-7}$ mbar. These three pictures were aligned with each other. A separate computer program was used to overlay the images, draw contours of the grains and calculate the size distribution. These are shown in Fig. 17 and Fig. 18 respectively.

Three tilt angles are used to ensure the visibility of all the grains. Sometimes it is sufficient to obtain grains for just one tilt angle.

## 5. TEM sample preparation.

Transmission electron microscopy is a technique for producing extremely high resolution images. As such, the TEM is very valuable for the analysis of VLSI devices [4]. Limitation of the application of this technique arise from the difficulty in preparing the samples sufficiently thin by grinding and broad beam ion milling. Recently developed techniques to prepare TEM samples on a FIB system has solved this problem. In principle, a TEM sample is prepared by cleaving a bar of 50 μm width and a length of 2 mm. out of a wafer or device. The feature of



15° tilt 20° tilt 25° tilt

Fig. 16. Channelling contrast of Al grains. Three tilt angles are used.

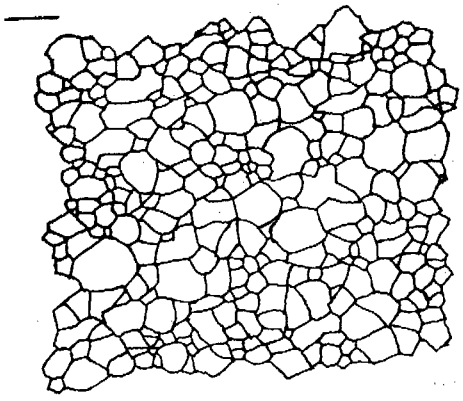


Fig. 17. Contours of the grains shown in Fig. 16.

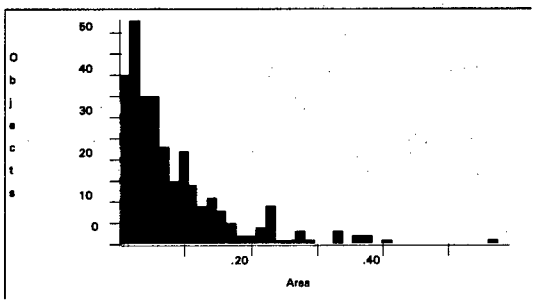



Fig. 18. Calculated grain size distribution.

interest must be in the centre of this bar. This bar is mounted on a TEM grid. This grid then is clamped in a TEM sample holder and brought into the sample chamber of the FIB.

After Pt deposition, on both sides of the deposited Pt line, bulk material is removed unto a depth of 10 μm. by using a beam current of 20 nA and enhanced etch. The membrane, left in the centre of the bar, is thinned and polished to a thickness of about 0.1μm with a beam current of 70 pA.

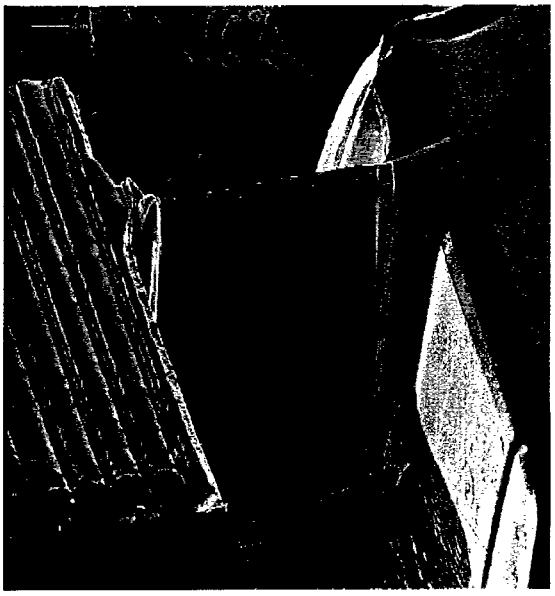


Fig. 19. FIB made TEM sample.

Fig. 19 shows a FIB made TEM sample of contact fingers of a transistor. In this particular case the TEM sample was made on a leakage spot observed by a Photo Emission Microscopy (PEM) See Fig. 20. The spot is indicated by a circle. The TEM images in Fig. 21 shows the cause of the problem: A severely damaged TEOS structure.

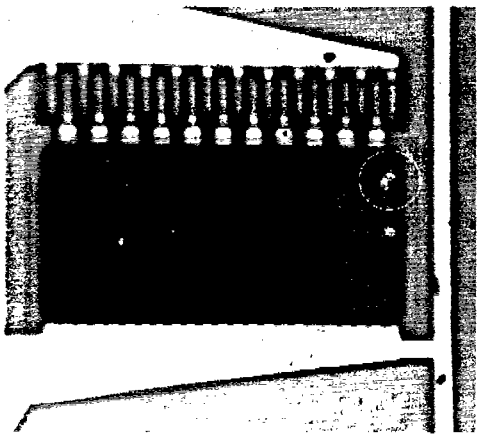


Fig. 20. PEM images of a localised leakage.

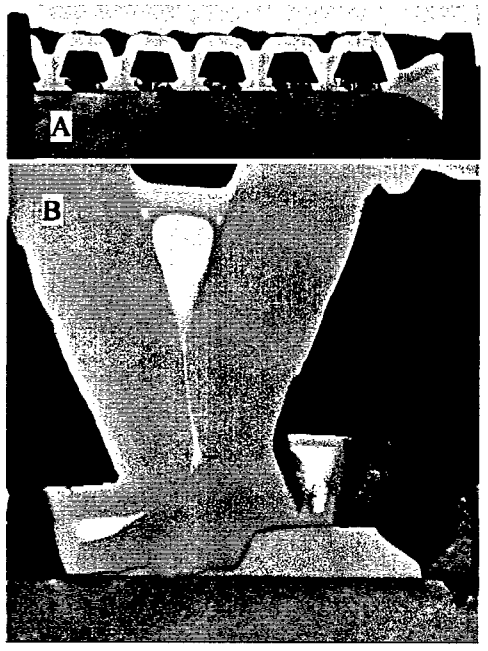



Fig. 21. TEM image. Image A on the top shows a overview of the contacts. Image B is a magnification of the area indicated by the arrow.

## 6. Conclusion.

In this article several examples have been shown of the use of the FIB in failure analysis and process monitoring in semiconductor industry. The use of the FIB as a scanning electron microscope provides us with a lot of information about the material under analysis. It should be noticed, however, that interpretation of the obtained images is not always straight forward. Because of the accuracy and speed to prepare samples for further analysis, the FIB has become an essential tool in a failure analysis laboratory.

## Acknowledgements

The author would like to thank Frits v.d. Elshout for the SIMS measurements, Ann de Veirman and Herman Peters for the TEM photographs. Bert Otterloo and Wim Claassen are thanked for valuable discussions and critically reading the manuscript.

## References

[1] Olson T.K., Lee R.G. and Morgan J.C. Contrast mechanisms in focused ion beam imaging. Proc. ISTFA 1992, pp 373 -382.

[2] Dingle T., Bickford C.U., Cook D.B., Do D.D. Gerlach R.L., Li J.Z. and Rasmussen J. The integration of a high performance quadruple SIMS facility with a $Ga^+$ LMIS based FIB instrument Proc. of the $10^{th}$ international conference on secondary mass spectrometry pp 517 - 520.

[3] Nikawa K. Applications of focused ion beam technique to failure analysis of very large scale integration: A review. J. Vac. Sci. Technol. B 9 (5) (1991) 2566 - 2577

[4] Pantel R., Auvert G., Mascarin G. Focused ion beam sample preparation, transmission electron microscopy and electron energy loss spectroscopy analysis of advanced CMOS silicon technology interconnections. Microelectronic Engineering 37/38 (1997) pp 49 -57.

PERGAMON

Microelectronics Reliability 38 (1998) 877–882

MICROELECTRONICS RELIABILITY

# Analysis of $I_{ddq}$ failures by spectral photon emission microscopy

M. Rasras, I. De Wolf, H. Bender, G. Groeseneken, H. E. Maes
S.Vanhaeverbeke*, P. De Pauw*.

*IMEC, Kapeldreef 75, B-3001 Leuven, Belgium*
** Alcatel-Microelectronics, Westerring 15. B-9700, Oudenaarde, Belgium*

**Abstract**

In this work, a fast identification of $I_{ddq}$ failures using spectroscopic photon emission microscopy (SPEMMI) is proposed. The spectra obtained from failure sites on the $I_{ddq}$ failed chips were compared with the ones of known defective components. Four distinguishable spectra categories were identified. They were attributed to gate oxide breakdown, metal shorts, blackbody radiation, and ESD caused junction spiking. The focused ion beam (FIB) technique was used to look at the damage sites for confirmation of the SPEMMI results.
© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Emission microscopy is commonly used for detection and localisation of failures in ULSI devices. In order to identify the failure modes, destructive and often time consuming techniques are used such as SEM, FIB, etc. Spectral analysis of the light emitted from failures of ULSI devices is a promising, fast alternative for failure identification. In fact, different defects, originating from different physical mechanisms, are expected to emit light with different spectral characteristics. Therefore, their spectra could be a unique means for finger print analysis of the defects [1-3].

The purpose of this paper is to show that Spectroscopic Photon Emission Microscopy (SPEMMI) allows fast identification of failures. Our SPEMMI system is based on a conventional Hypervision photon emission microscope (EMMI) equipped with a highly sensitive detector (GENIII-NIR). The EMMI system was modified to allow for continuous wavelength analysis (SPEMMI) with high resolution and sensitivity, using a prism/lens spectrometer [4]. Spectral analysis can be done by lifting the detector automatically with the aid of a pneumatic mechanical system, and by then inserting the prism/lens between the detector and the optical microscope [5, 6].

To demonstrate this failure analysis technique, various types of $I_{ddq}$ failed devices were analysed with the SPEMMI system. Spectra obtained from different failures were recorded and compared with spectra from known defective components. In addition, for further confirmation of the SPEMMI results, the focused ion beam (FIB) was used to look in detail at the damage location.

## 2. Emission from the $I_{ddq}$ failure sites

The $I_{ddq}$ failed chips were biased at 5 V and EMMI was used to localise the failure. To obtain more information on the defective component, the location of the light emission on the chip was compared with the chip lay-out. The lay-out showed that failures were found at different locations: on top or at the edge of connected as well as not-connected gates of nMOS and pMOS transistors, on top of W-plugs, and near metal lines as shown in fig. 1a-d, respectively. However, this lay-out information was not conclusive for the defect identification.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00106-1

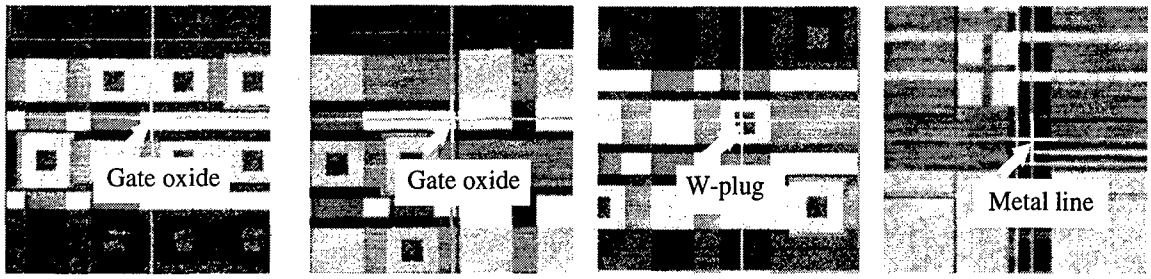



Fig .1. Lay-out images showing the location of failures at different positions on the chip. The failures were localised by EMMI, and their positions are indicated by the arrows. White: Poly (gate oxide). Dark squares: W-plugs. White cross: the location of the failure. Grey: Metal lines.

Spectral analysis of the light emitted by these defects revealed four different categories of spectra as shown in Fig. 2.a-d, and they are corresponding to the failures indicated in Fig. 1.a-c, respectively. The first spectrum (type A, Fig. 2a.) has a broad, and flat intensity distribution. The emission intensity of the second spectrum category, Fig. 2.b, type B, covers also a wide wavelengh range, but it has a lower intensity at shorter wavelength compared to the type A spectrum. By referring to the lay-out, both failures were found to occur mostly on top, or at the edge of gate oxides, indicating that the failures are due to probably a gate oxide breakdown. Spectra of the third category Fig. 2.c, type C, are characterized by a strong, and very narrow emission intensity only emitted at the long wavelength range with a cut at about 800 nm. This type of spectrum was found between a W-plug and a gate oxide. The last case, type D, corresponds to a faint emission. In most of the failed devices belonging to this class, the location of the failures did occur near metal lines or W-plugs.

In order to have a better understanding of the physical nature of these failures, different emission

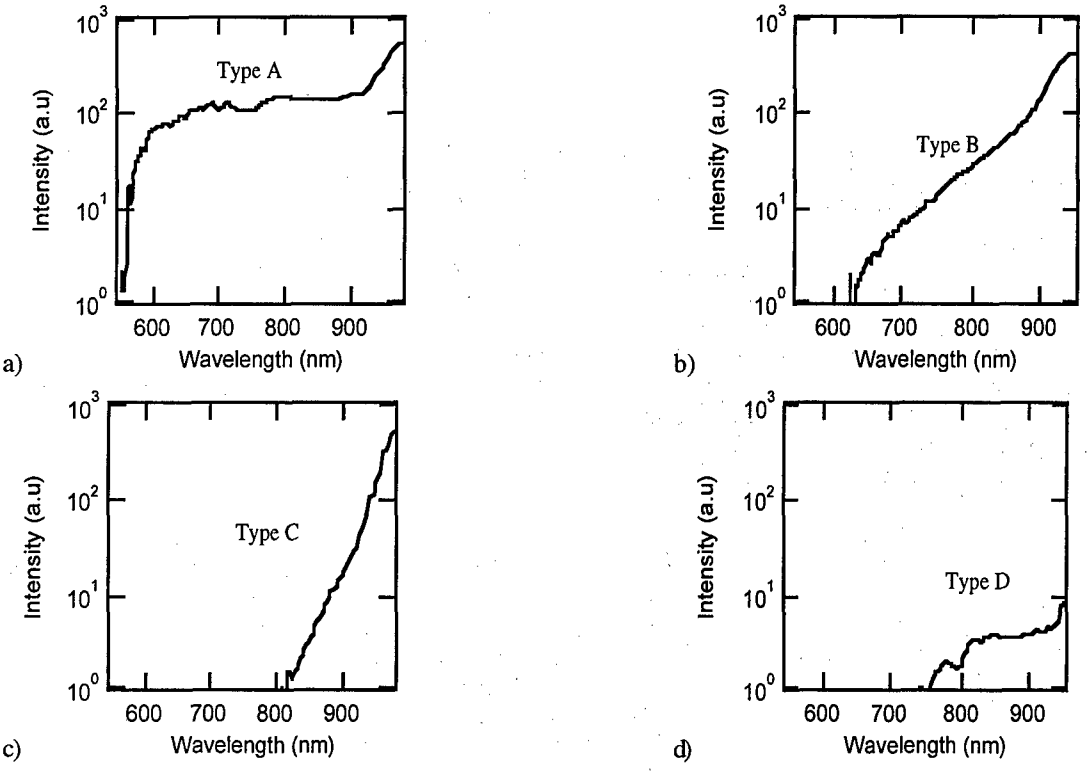



Fig. 2. Emission spectra from the failures: a) A-Type b) B-Type c) C-Type, d) D-Type.

spectra from several single devices were investigated. Emission from gate oxide capacitor breakdown and from a pn junction in the forward and reverse biasing conditions were studied. In addition, light emitted from failures due to ESD spiking was investigated. Emission due to blackbody radiation was investigated as well. Spectra collected from all these devices were compared to the ones obtained from the failures on the chips in order to have a conclusive identification for each failure type.

## 3. Spectra from different emission mechanisms

### *3.1 pn Junction*

Emission spectra from a forward and reverse biased diode were measured at different biasing currents. Fig. 3 shows that increasing the biasing current does not affect the shape of the forward nor reverse spectrum. In the reverse biased diode, the current originates from carrier multiplication in avalanching conditions. These carriers have a wide energy distribution. Therefore, the light emission is believed to originate from intraband transitions or bremsstrahlung radiation, which both yielding a wide spectral light distribution [3]. The spectrum generated in the avalanching process has a broad and flat intensity distribution similar to the failure of type A. However, to obtain avalanching, a larger bias voltage than the 5V which was applied to these chips is needed. This makes this emission mechanism unlikely to be the origin of the type A failure.

In the forward bias case, the spectrum is characterized by a narrow emission range at long wavelengths. In fact, this emission is generated by conduction-to-valence band recombination, and its spectrum shifts toward shorter wavelengths and higher intensities as the biasing currents increase. This spectrum is similar to the emission of a failure of type C.

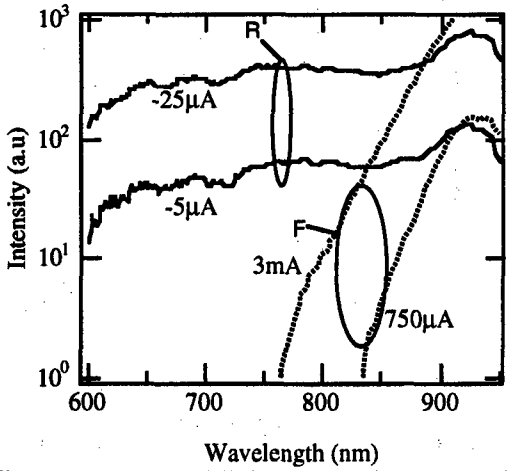



Fig. 3. Spectrum of light emitted from forward and reversed bias diode at different current densities.

### *3.2 Gate oxide breakdown*

Dielectric breakdown happens when a conduction path occurs between the substrate and the gate contact of the capacitor. After breakdown, local light emission can be detected at the breakdown location. In order to study the corresponding spectrum, a p-type capacitor ($n^+$ poly / p-substrate), with oxide thickness of 8.5 nm, was stressed until it was broken, using negative constant current stress (-32 μA). The electric field across the oxide was about 10MV/cm. After breakdown, the spectrum from the emission spot was immediately recorded, Fig. 4a. This spectrum is relatively flat and shows a broad distribution. It is similar to the failure spectrum of type A.

When the voltage applied to the failed capacitor was increased, we noticed that the breakdown path could be further degraded. The leakage current through it increased. In addition, the spectral characteristics of the emitted light changed as can be seen in Fig. 4.b. For larger voltages and higher leakage currents, the spectrum becomes narrower. For a complete breakdown, with low resistance, the spectrum characteristic is similar to that of a forward biased diode as shown in Fig. 3. At this stage, the

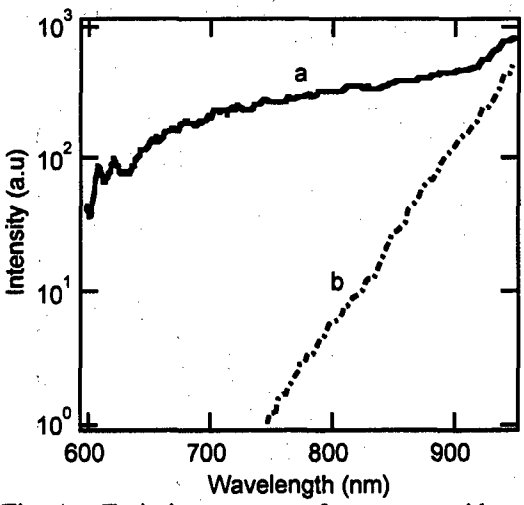



Fig. 4. Emission spectra from an oxide after breakdown, recorded at different voltages. a) Immediately after breakdown V=-4V, I=-414.5μA. b) Completely damaged V=-3V, I=-4.1mA

breakdown path behaves as a forward biased diode. This shows that the spectrum from oxide breakdown can also be of type C.

### 3.3. Blackbody radiation

The EMMI detector is not very sensitive to thermal radiation, because its sensitivity is limited between 500 and 1000 nm. However, for temperatures higher than 350 °C, the quantum efficiency of the detector is high enough to detect radiation at the long wavelength range ($\lambda$>850 nm). As a result, if emission originates from heat, it will show up as a very faint light in the SPEMMI at the long wavelength side of the spectrum. Therefore, spectra of type C can be related to heating. This heating can occur when a high current passes through a high resistive path.

Fig. 5 shows a spectrum emitted by a poly-Si line, biased at 27V. It is compared with the emission from a spiral lamp at a temperature of 570 °C. Similar measurements indicated that emission can be detected for temperatures higher than 350 °C. The temperature of the blackbody radiator can be determined by fitting the Planck equation to the spectrum.

### 3.4. ESD

Light emitted from a grounded gate nMOS (ggmos) was studied in avalanching conditions, and after the device was damaged by ESD pulses. The spectrum was recorded at three different reverse biasing currents for a fresh device in the avalanching regions: (-50 μA, -12.1V), (-100 μA, -12.3V), (-150 μA, -12.4V) as shown in Fig. 6a, b, and c,

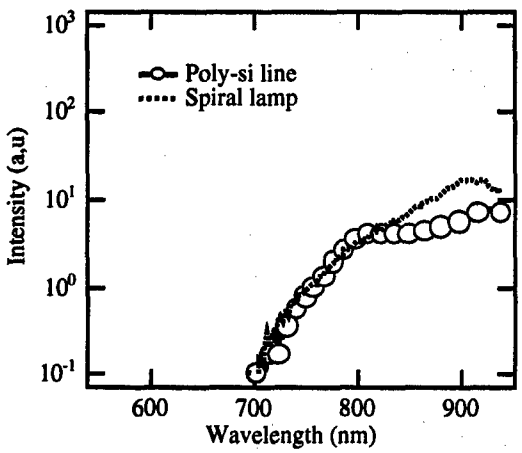



Fig. 5 Emission from a poly-Si line biased at 27V, compared to the emission from a spiral lamp.

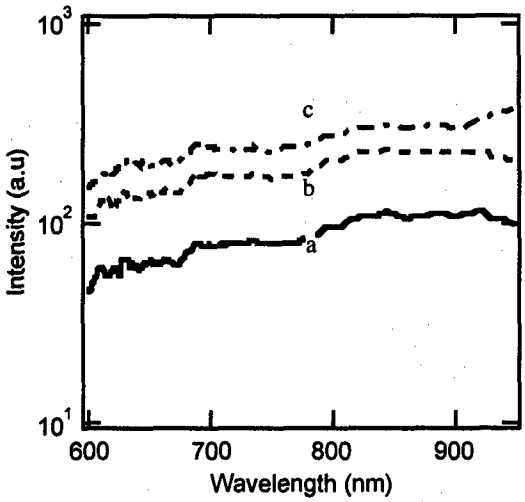



Fig. 6. Emission spectra from ggmos device at different reverse biasing conditions. a) -50 μA, -12.1V, b) -100 μA, -12.3V, c) -150 μA, -12.4V

respectively. These spectra originate from the same mechanism as the reverse biased diode, and they have a similar intensity distribution. As the current increases, the light intensity increases, and the spectrum shifts to higher intensities without changing its spectral shape. In fact, these curves are similar to the curves of type A failures. However, avalanche is not expected in the chip because emission is already observed at 5V, which is too low to give avalanche.

The device was degraded with ESD pulses until it was damaged. This damage was in similar devices found to be a filamenting path between the drain and

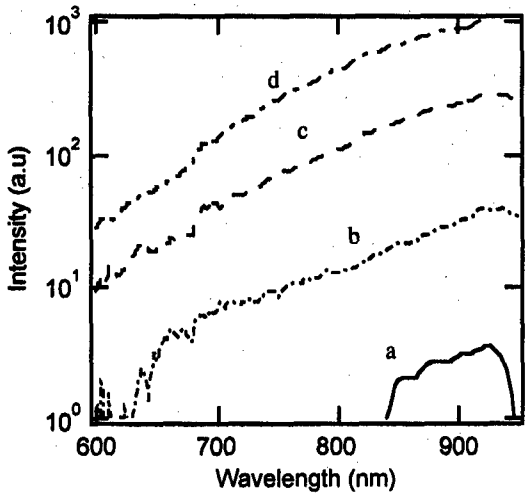



Fig. 7. Emission from failed ggmos a) 30 mA, 2.3 V. b). 40 mA, 3.0 V. c) 50 mA, 3.7 V. d) 40 mA, 4.2 V.

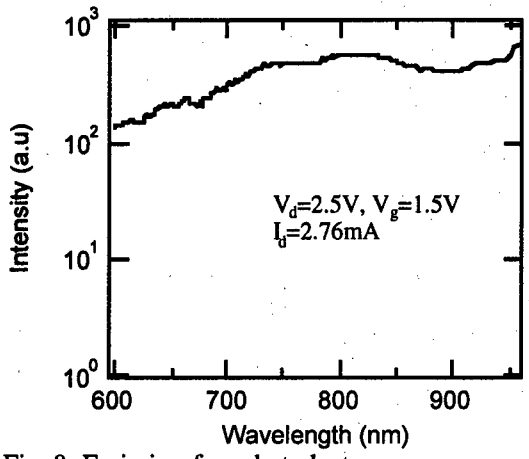



Fig. 8. Emission from hot electrons.

the source. Spectra of light emitted from the failed device are shown in Fig. 7 at different applied currents. Comparison of Fig. 6 and Fig. 7 shows that damaging the devices affects the shape of the spectra, i.e. the spectra become steeper.

The emission mechanism for this failure is still not fully understood, it could be attributed to light emission from hot electrons, or heating at the location of the damage. Spectra emitted from these damaged devices were compared to those due to hot electrons in an nMOS device as shown in Fig. 8. The spectrum of hot electrons differs from the one of the defect which indicates that hot electrons are not the reason for the light emission. On the other hand, the shift in intensities toward shorter wavelengths for large currents, as seen in Fig. 7a, b, c, d, indicates that the emission could be due to thermal heating. A blackbody radiation model was used to fit these spectra, resulting in temperatures of 655 °C (30 mA), 835 °C (40 mA), 1030 °C (50 mA), and 1195 °C (60 mA), respectively. The accuracy of these temperature data is within 100 °C. Fig. 9 shows a linear relation between the temperature and the fourth root of the power dissipated in the device (I.V). This supports the idea that the emission is due to temperature, because the power is related to the temperature T by:

$$P = \sigma * T^4 \quad (1)$$

where $\sigma$ is the Stefan-Boltzmann constant, and P is the power dissipated in the device. This type of emission has similar characteristics as that of failure type-B. Therefore, these failures are probably due to blackbody radiation.

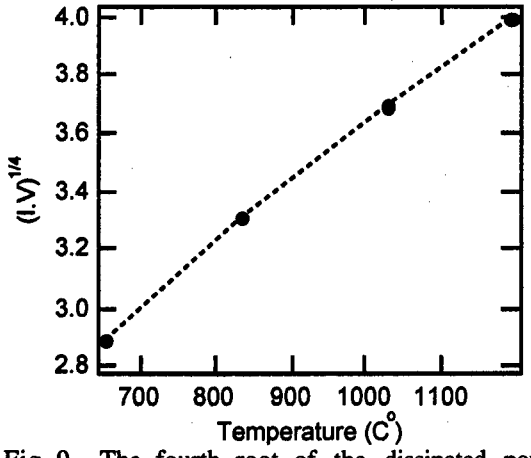



Fig. 9. The fourth root of the dissipated power $(I.V)^4$ Vs temperature.

### 3.5. FIB

For further confirmation of these spectral results, the FIB technique was used to investigate the damaged $I_{ddq}$ device at the location of the light emission. For the devices types A, B, C which were investigated with FIB, no special damage features could be identified. However, for the D-type, a contact between two metals was detected, as shown in Fig. 10a and b for two devices. The FIB results confirm that the emission in this case is due to heat, originating in a high resistive path between the two metal lines.

## 4. Discussion

Table I summarises the correlation between the spectra obtained from the $I_{ddq}$ failed devices, and the possible corresponding failure mechanism. For type A, the failure originates from a softly damaged gate oxide, or from a reversed biased junction. Spectra of type B can be correlated to blackbody radiation in a filament between the drain and the source of a transistor, or ESD induced damage. The third

Table I.
Spectra failure categories, and their corresponding failure mechanisms.

| failure type | possible failure mechanism. |
|---|---|
| A | gate oxide breakdown, reverse biased pn junction |
| B | ESD filament (heath, spike), gate oxide breakdown, or heat |
| C | gate oxide breakdown, forward pn junction |
| D | shorts between metal lines, heath. |

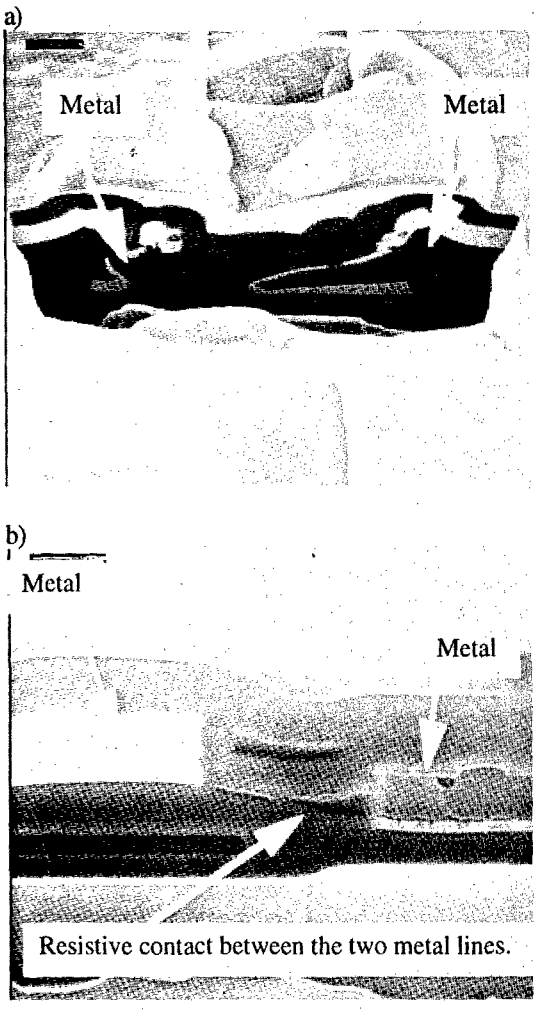



Fig. 10a, b: FIB images showing a failure due to a contact between two metal lines

spectrum of type C can be due either to a forward diode, or to a severely damaged gate oxide. In the last case, type D, it is clear that the emission is due to heat radiation, as it was verified by the FIB experiments.

## 5. Conclusion

In this paper, an attempt was made to correlate the emission spectra from failure sites of $I_{ddq}$ failed devices, and the spectra from known failure mechanisms. Three failure mechanisms were identified: metal shorts (heat), gate oxide breakdown, and ESD junction spiking. This correlation was confirmed using FIB for the case of metal shorts. Further confirmation is still required for the other cases by using SEM combined with reactive ion etching (RIE) for delayering. These first results indicate that the SPEMMI technique in combination with a comparison of the failure position with the lay-out, can be used as a fast failure identification tool, that can give non-destructive information on the failure cause.

## Acknowledgements

This project was sponsored by the Flemish IWT960107 project (KWINT). Simac Masic b.v. (Veldhoven, The Netherlands) is acknowledged for help in design and fabrication of the mechanical parts of the SPEMMI. Dr. K. Bock, and C. Ross are acknowledged for providing the ESD devices, and for their helpful discussions.

## References

1. Kölzer et al. J. Appl. Phys. 71 (11), pp. R23-R41, 1992.
2. Tao J. M, W. K. Chim, D. S. H. Chan, J. C. H. Phang, and Y. Y. Liu. Proc IRPS, pp. 360-365, 1996.
3. Bude J., N. Sano and A .Yoshii. Phys. Rev. B, **45** (11), pp. 5848, 1992.
4. K. de Kort and P. Damink. Proc. ESREF, pp. 45-52, 1990.
5. M. Rasras, I. De Wolf, G. Groeseneken, H. E. Maes, S. Vanhaeverbeke, P. De Pauw. Proc. ISTFA Symp. 1997, pp. 153-157, Santa Clara, USA.
6. M. Rasras, I. De Wolf, G. Groeseneken, H. E. Maes. Microelectronics and reliability. Vol. 37: pp. 1595-1598; 1997.

PERGAMON

Microelectronics Reliability 38 (1998) 883–888

MICROELECTRONICS RELIABILITY

# Near-field-optical-probe induced resistance-change-detection (NF-OBIRCH) method for identifying defects in Al and TiSi interconnects

K. Nikawa[a], T. Saiki[b], S. Inoue[c], M.Ohtsu[b,d]

[a] *NEC, Kawasaki, JAPAN*
[b] *Kanagawa Academy of Science and Technology (KAST), Kawasaki, JAPAN*
[c] *T.D.I., Co.,Ltd. (Technology Development of Information processing), Kawasaki, JAPAN*
[d] *Tokyo Institute of Technology, Yokohama, JAPAN.*

**Abstract**

The optical-beam-induced resistance-change-detection (OBIRCH) method has been improved by using a near-field optical probe as the heat source instead of a laser beam. The resulting NF-OBIRCH method has two advantages over the conventional OBIRCH method.

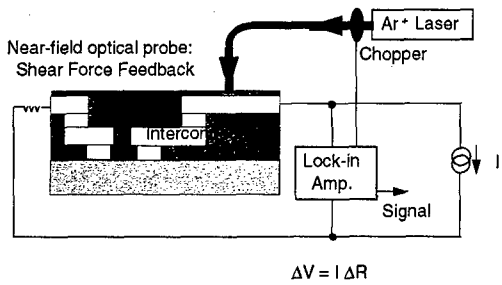
(1) Its spatial resolution is higher.

(2) The optical-beam-induced resistance change caused by heating can be observed when the aperture size is zero without interference from the photo current caused by electron-hole-pair generation. In the conventional OBIRCH method, the laser beam creates not only the resistance change, but also the photo current that can mask the resistance change signals. © 1998 Elsevier Science Ltd. All rights reserved.

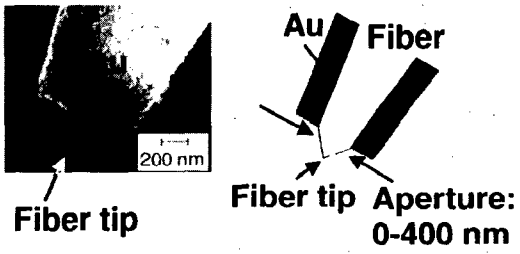
## 1. Introduction

The identification of voids and Si nodules in the Al stripes of integrated-circuit-device chips is a key part of failure analysis and process monitoring in the semiconductor industry. Several approaches to identifying voids have been proposed, including thermal-wave modulated optical reflectance imaging [1], backscattered electron imaging using field-emission scanning electron microscopy [2], and ultrahigh-voltage transmission electron microscopy [3]. The optical-beam-induced resistance-change-detection (OBIRCH) method (first proposed by Nikawa and Tozaki [4]) has been shown to be, in some respects, more useful in void detection than these other methods. It can identify Si precipitates in metal lines [5], voids under vias [6], high resistance areas at the bottoms of vias, and high resistance sections in TiSi lines [7]. Moreover, it can localize current paths on VLSI chips [8].

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00100-0




(a) Setup




(b) Near-field optical probe

Fig. 1. Experimental setup

The OBIRCH method uses two main simultaneous processes: laser-beam heating and resistance-change detection. Since it was first introduced it has been modified and improved in various ways. The sensitivity of its resistance-change detection has been improved by reducing the sample temperature during measurement [9], by increasing the laser power [6,7], and by improving the technique used for detecting the resistance changes [7,10,11]. The wavelength of the laser used for heating has been changed from 633 to 1300 nm [8, 10, 12]. This change has made it possible to heat the Al stripes on the Si chips from the backside of the chips, and to reduce the intensity of the optical-beam induced current (OBIC) , thereby no longer masking the OBIRCH signal.

We have modified the OBIRCH method to use a near-field optical probe [13] as the heat source instead of a laser beam. Testing of this NF-OBIRCH method showed that it is superior to the conventional one.

## 2. Experiment

### 2.1. Experimental setup

Figure 1(a) shows the experimental setup of NF-OBIRCH. To test our NF-OBIRCH method, we introduced an Ar laser (a mixture of 488 nm and 514.5 nm) into a near-field optical probe. Figure 1(b) shows a scanning electron microscope (SEM) image and an illustration of a typical near-field optical probe we used. The apex aperture sizes of the probe were changed from 0 to 400 nm. The probe was scanned in the area of interest, using a shear force feedback technique which are usually utilized in a typical near-field optical microscope system. The conventional amplifier system [5], which uses a constant voltage source and a specially designed amplifier, did not amplify the resistance changes well because the resistance changes were much weaker than those in the conventional OBIRCH method. We used a constant current source instead of the constant voltage source . For amplifying the resistance change signals, we used a combination of a modulated-laser-beam and a lock-in amplifier instead of the specially designed amplifier. The voltage changes during scanning are introduced into an external input terminal of a typical scanning probe microscope system so as to produce an image of resistance changes.

The voltage change is simply described as:

$$\Delta V = I\,\Delta R \qquad (1)$$

where $\Delta V$ is a voltage change when a laser beam is irradiated, $I$ is a current applied by a constant current source, and $\Delta R$ is a resistance change when a laser beam is irradiated. From the term of $I$, current paths can be observed [8], and from the term of $\Delta R$, various defects can be observed [5].

### 2.2. Observed results

The samples were 0.3-um-wide Al lines and 0.2-

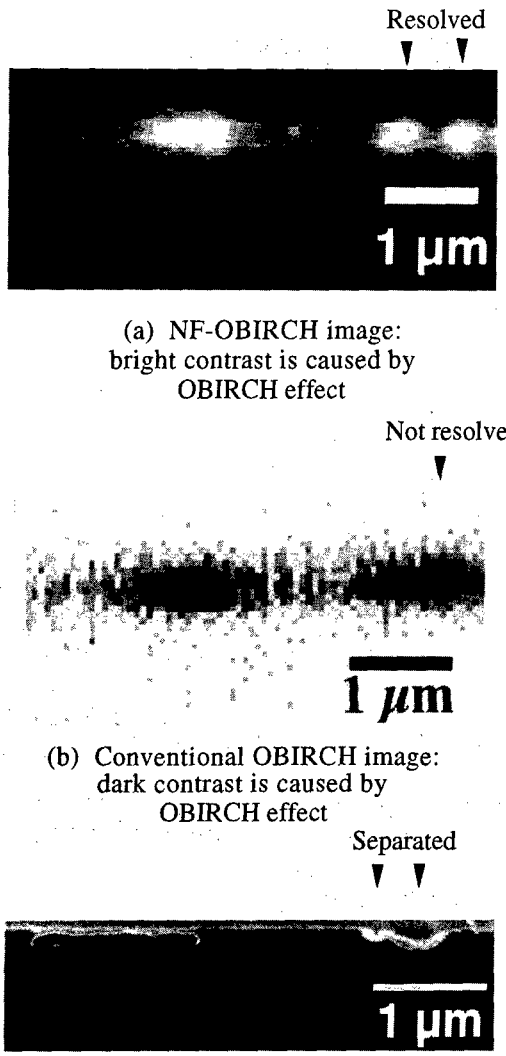



Fig. 2. 0.3 μm wide Al line

um-wide TiSi lines fabricated using processes for VLSI development.

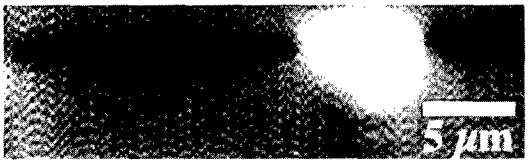
*2.2.1. Al line*

Figure 2 shows the results of Al line observations. Figure 2(a) shows the NF-OBIRCH image where the apex aperture size of the probe was 400 nm. Figure 2(b) shows the conventional OBIRCH image using 633 nm laser. The voids identified with the NF-OBIRCH method and the conventional OBIRCH method were confirmed to exist by cross-sectional scanning ion microscopy (X-SIM). Figure 2(c) shows the X-SIM image. Two voids can be seen in the SIM image. A left hand side one is larger than a right hand side, and the right hand side void are separated into two parts. The left hand side void are clearly shown both in the NF-OBIRCH image (Fig. 2(a)) and in the conventional OBIRCH image (Fig. 2(b)). The separation of right hand side void are resolved in the NF-OBIRCH image (Fig. 2(a)) but not resolved in the conventional OBIRCH image (Fig. 2(b)). These results shows that the spatial resolution of voids identification is better with the NF-OBIRCH method than with the conventional one.

*2.2.2. TiSi line*

TiSi lines whose resistance was higher than normal were first observed using the conventional OBIRCH methods: using both infrared OBIRCH (IR-OBIRCH: 1300 nm) and visible light OBIRCH (VL-OBIRCH: 633 nm). The results which were already reported in ref. [7] are shown in Fig.3. Figure 3 shows that the spatial resolution with VL-OBIRCH image (Fig.3(b)) is better than with IR-OBIRCH image (Fig.3(a)) . Both images show dark and bright contrasts. The dark contrasts correspond to the OBIRCH effect at a normal TiSi line. The origin of bright contrasts must be the OBIRCH effect (resistance change caused by thermal effect: negative temperature coefficient of resistance causes bright contrast) or an OBIC effect (photo current caused by electron-hole pair generations), but had not been able to be confirmed [7].

The NF-OBIRCH method can cause only the OBIRCH effect when no aperture is present (metallized probe). We used the metallized probe shown in Fig. 4(b). Figure 4(a) shows a NF-OBIRCH observation result using the metallized probe. The bright contrast in Fig.4(a) corresponds to the OBIRCH effect: note that bright and dark contrasts are inversed compared to the conventional OBIRCH images because




(a) IR-OBIRCH image: 1300 nm laser.

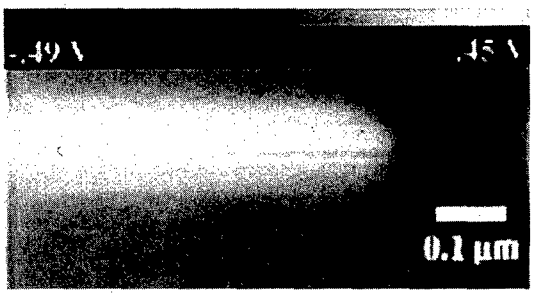


(b) VL-OBIRCH image: 633 nm laser.

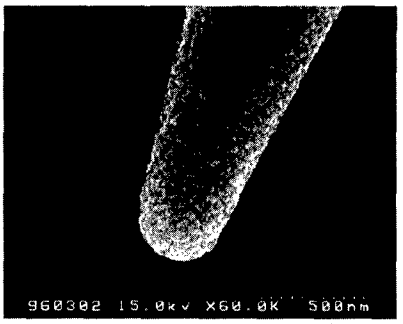
Fig. 3. Conventional OBIRCH images of a 0.2 μm wide Tisi line: dark contrasts are caused by OBIRCH, the origin of bright contrasts was not able to be identified in ref. [7].




(a) NF-OBIRCH image: metallized probe. bright contrast is caused by OBIRCH effect no OBIC contrast is appeared

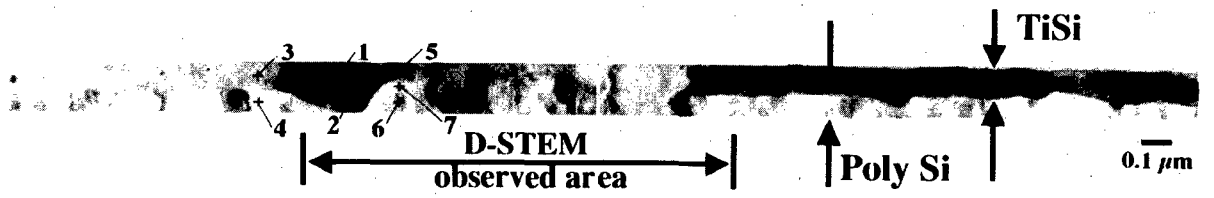



(c) NF-OBIRCH image: apertured probe. bright contrast is caused by OBIRCH effect dark contrast is caused by OBIC effect

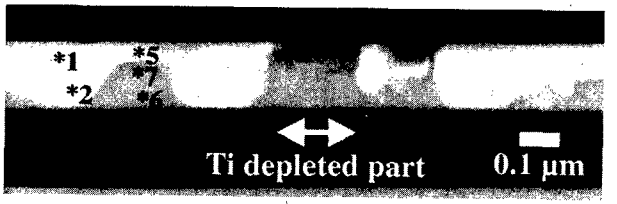



(b) SEM image of near-field optical probe: metallized probe.

Fig. 4. NF-OBIRCH images of a 0.2 μm wide Tisi line




(a) Cross sectional TEM image




(b) Cross sectional D-STEM image: EDX results at 1,2,5-7 showed bright area is TiSi and gray area is only Si

Fig. 5. TEM images of a 0.2 μm wide TiSi line.

the resistance increase corresponds to the increase of the signal (voltage) in the NF-OBIRCH method, but it corresponds to the decrease of the signal (current) in the conventional OBIRCH method. No contrast (0.00 V) is observed in Fig. 4(a) at the part where the bright contrasts are observed in Figs. 3(a) and 3(b). This result suggests that the bright contrasts shown in Figs. 3(a) and 3(b) were cause by the OBIC effects.

The NF-OBIRCH method can cause both OBIRCH and OBIC effects when the aperture is present (apertured probe). We acquired the same area NF-OBIRCH image with the apertured probe as with the metallized probe. Figure 4(c) shows an apertured-probe-NF-OBIRCH observation result. The bright contrast in Fig.4(c) corresponds to the OBIRCH effect and the dark contrast must correspond to the OBIC effect. The signal of the dark contrast is as strong as that of a bright contrast (-0.49 V and 0.45 V). This result must have confirmed that the bright contrasts shown in Figs. 3(a) and 3(b) were cause by the OBIC effects.

In order to confirm the physical cause of the contrasts, observations using cross sectional transmission electron microscopy (X-TEM) have been conducted. Figure 5(a) shows the X-TEM image. The right hand side shows a normal structure where an upper half is TiSi and a lower half is Poly Si. A structure of middle part shown as "D-STEM observed area", on the other hand, is not clear. In order to image this part clearly, dark field scanning TEM (D-STEM) observation has been conducted. Figure 5(b) shows the D-STEM image. White areas correspond to heavier atoms and gray areas correspond to lighter atoms. In order to identify the elements, the points numbered from 1 to 7 in Figs. 5(a) and 5(b) are analyzed by energy dispersive X-ray analyzer (EDX). The results showed Ti and Si atoms exist at the white areas in Fig. 5(b) and only Si atoms exist at the gray areas in Fig. 5(b). Figure 5(b) shows that the gray area corresponding only Si, shown as "Ti depleted part", exist across the TiSi line. This must be the cause of high resistance of this line.

## 3. Discussions

The exact mechanism of the OBIC effect at Ti depleted part must be discussed. Why the bright contrast corresponding to the OBIC effect is observed in

Fig. 3(a), when 1300 nm laser is utilized, is left to be answered. In the conventional OBIRCH method, the 1300 nm laser is usually utilized in order not to produce OBIC [8, 10, 12], because the energy of 1300 nm laser (0.95 eV) is less than Si band gap energy (1.12 eV). The key to solve this puzzle is the Ti detection limit of EDX. The detection limit of Ti is about 1 %. Ti may be exists enough to produce impurity levels. The impurity level of Ti is 0.21 eV under conduction band: 0.91 eV from valence band. The 1300 nm laser, therefore, can generate electron-hole pairs via Ti-levels.

## 4. Conclusions

We have improved the optical-beam-induced resistance-change-detection (OBIRCH) method by using a near-field-optical-probe as the heat source instead of a laser beam. The resulting NF-OBIRCH method has two advantages over the conventional OBIRCH method.

(1) Its spatial resolution is higher.

(2) The optical-beam-induced resistance change caused by heating can be observed when the aperture size is zero (metallized probe) without interference from the optical beam induced current (OBIC) : In the conventional OBIRCH method, the laser beam creates not only the resistance change, but also an OBIC that can mask the change.

## Acknowledgements

We thank H. Iida, E. Inuzuka, H. Terada, Y. Nakajima, and H. Suzuki of Hamamatsu Photonics for their very valuable technical support of the IR-OBIRCH method we used; we also thank T. Fukase, H. Tsuda, and Y. Yamada for supplying the samples and for their technical suggestions.

## References

[1] Smith W et al., Proc. Int. Rel. Phys. Symp., IEEE, p. 200(1990).

[2] Besser P et al., J. Appl. Phys. vol. 72 p. 3792(1992).

[3] Tanikawa A et al., Proc. Int. Rel. Phys. Symp., IEEE, p. 209(1990).

[4] Nikawa K and Tozaki S, Proc. Int. Symp. Testing and Failure Analysis, ASM International, pp. 303-310(1993).

[5] Nikawa K et al. Jpn. J. Appl. Phys., vol. 34, Part 1, no. 5, pp. 2260-2265 (1995).

[6] Nikawa K and Inoue S, Proc. ESREF, p. 307-312(1995).

[7] Nikawa K and Inoue S, Seventh Int'l Conf. Defect Recognition and Image Processing in Semiconductors (DRIP VII), Germany, Sept. 1997, Proc. DRIP VII, Inst. Phys. Pub., Bristol, UK, pp.37-46 (1998).

[8] Nikawa K and Inoue S, Proc. Int. Rel. Phys. Symp., IEEE, pp. 346-354 (1996).

[9] Kawamura N et al. IEICE Trans. Electron., vol. E77-C, no. 4, pp. 579-584 (1994).

[10] Cole Jr. E I et al. Proc. Int. Rel. Phys. Symp., IEEE, pp. 129-136 (1998).

[11] Wen Q and Clarke D R. Appl. Phys. Lett. vol. 72, no. 15 (13 April 1998), pp. 1920-1922: This paper describes a modification of the OBIRCH method, but does not refer to papers related to the OBIRCH method.

[12] Nikawa K and Inoue S, Proc. Int. Symp. Testing and Failure Analysis, ASM International, pp. 387-392 (1996).

[13] Saiki T et al. Appl. Phys. Lett. vol. 68 pp. 2612-2614 (1996).

PERGAMON

Microelectronics Reliability 38 (1998) 889–893

MICROELECTRONICS RELIABILITY

# A new adaptive amplifier for biased electron beam induced current applications

M. Ciappa[a], P. Malberti[a], P. Furcas[c], and M. Vanzi[b]

[a] *Swiss Federal Institute of Technology (ETH), Integrated Systems Laboratory, ETH-Zentrum, CH-8092 Zurich*
[c] *University of Cagliari, DIEE-INFM., Piazza d'Armi, I-09123 Cagliari*
[b] *University of Cagliari, Physics Dept., Via Porcel, I-09123 Cagliari*

**Abstract**

A novel amplifier for Electron Beam Induced Current (EBIC) mapping is presented which enables sample polarization and which provides at the same time good characteristics in terms of sensitivity, input bandwidth and noise-immunity. The proposed circuitry includes an adaptive self-compensation of the DC level due to the leakage currents. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

At present, Electron Beam Induced Current (EBIC) is still a valid technique for the characterization and the failure analysis of semiconductor devices. Basically, EBIC uses electrons injected by the primary beam of a scanning electron microscope (SEM) to produce a map of the local recombination efficiency within the semiconductor. The physical background of signal generation and carrier transport in space charge regions (junctions, Schottky contacts) under irradiation is already well understood since the 70's [1,5]. Briefly, the energy of the primary electrons impinging onto a probe is mainly dissipated by generation of phonons, photons, secondary electrons, and electron-hole pairs. In absence of a local electric field, electron-hole pairs rapidly recombine such that no EBIC current is generated and consequently no EBIC current is collected by an external amplifier. In converse, if the pairs are generated inside a space charge region, the intrinsic field separate the carriers in opposite directions, leading finally to an EBIC signal to be amplified by the external amplifier. The number of collected EBIC electrons is typically three orders of magnitude greater than the primary current, and under certain conditions, a locally reduced EBIC signal indicates the presence of recombination centers within the depletion zone (crystal defects, different doping concentration, contaminants, etc.). Usually, local recombination centers cause a modulation of the (pure) EBIC signal in the 0.5 to 5% range.

In general, the application of a reverse bias is not necessary for producing a reasonable EBIC map. However, a reverse bias can improve the collection of EBIC charges by expanding the space charge region, especially in those cases where just a weak EBIC signal is produced, like in micrometer scale structures or in device microsections where surface recombination is usually the dominating (parasitic) effect. Furthermore, reverse bias (polarization) is mandatory when integrated circuits are investigated for floating gate structures (MOS), latching sites ($I_{dd}$-monitoring), etc..

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00078-X

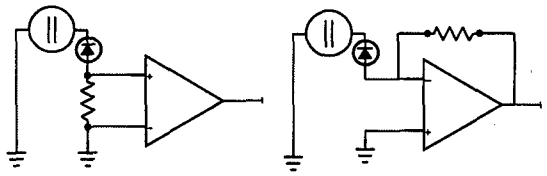


Fig.1 DC-coupled amplifiers

By applying a reverse bias, the fluctuations of the EBIC signal are superimposed either to the leakage current through the junction (large in the case of damaged devices) or to the supply current of the device under investigation. This yields a current offset and an increase of the shot noise which highly degrade the signal-to-noise ratio. Since these parasitic components are usually slow functions of time, they can virtually be suppressed by a capacitively-coupled amplifier, like that schematically represented in Fig.2 [2]. However, in practical applications such a simple circuitry results into the

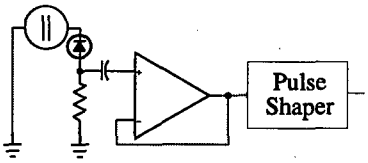



Fig.2 AC-coupled amplifier

loss of low frequency EBIC signal, as a consequence of the fact that the input-bandwidth of the amplifier is difficult to be matched with the input signal. This problem can be only partially solved by the DC-coupled amplifiers of Fig.1 or of [3]. In effect, in the case of damaged junctions, microsections, etc., such amplifiers are driven in saturation just by the large current offsets, which can be up to a decade larger than the pure EBIC signal. Even a manual correction of the offset as proposed in [4,6] is useless when the DC-component of the signal changes either with time or due to the primary beam conditions (acceleration voltage, probe current).

## 2. Experimental

The general specifications imposed to the new EBIC amplifier appear evident from the review: DC-coupled amplifier with good DC-signal rejection properties, self-compensation of the DC-component of the signal, reverse bias capability of the sample up to 12V, high sensitivity, and cut-off frequency of about 300 kHz at usual operating conditions. The amplifier which has been synthetized is represented in Fig.3, Fig.4, and Fig.5. It can be divided functionally into three stages: polarization-preamplifier, DC-filter with self-compensation, and impedance transformer.

### *2.1 Polarization-preamplifier*

The preamplifier is constituted by two (complementary) bipolar transistors TR1 and TR2. The base voltage of TR2 is controlled by the voltage follower OA1 which generates at the same time the reverse bias (up to 12V) directly applied to the base of TR1. Thus TR1 acts as an EBIC current amplifier with an amplification factor in the 50-300 range depending on the polarization, while TR2 is connected in common-base configuration.

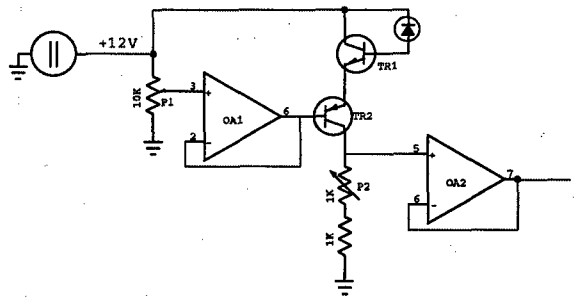



Fig.3 Polarization and preamplifier

The current-to-voltage conversion is performed by the series potentiometer P2. This results in a preamplifier sensitivity of $10^5$-3 $10^6$ V/A. Since present amplifier operates in the biased mode, the actual sensitivity is strongly improved by the increase in the collection efficiency due to the reverse polarization. The improvement in the sensitivity can be estimated [4] in two up to three orders of magnitude. It should be noted that this level of sensitivity is obtained by using a resistor in the low kΩ range. This turns in a significant reduction of the noise. Additionally, the bipolar transistor pair can be mounted in the immediate vicinity of the sample (in the vacuum chamber), thus avoiding the coupling of external noise through the wiring cables. For more critical applications, the sample and the bipolar transistor pairs can be cooled down to liquid nitrogen temperature. The

response of the preamplifier is linear for bias voltages below 5V. For higher bias voltages, the preamplifier behaves like a fast discriminator, delivering trapezoidal pulses with the rise- and fall-edges controlled by the response bandwidth of the system.

*2.2 DC-filter with self-compensation.*

The filter consists basically of a differential stage (OA5). The voltage output of the preamplifier is sensed by a high-input impedance buffer (OA2). Consequently, the output signal is splitted in two branches which are driven into the inverting and non-inverting inputs of OA5, respectively.

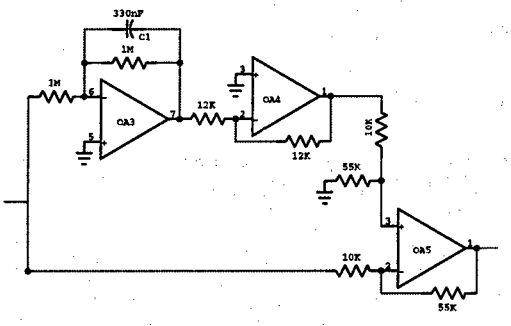



Fig.4 Filter with self-compensation

The function of the differential stage is to subtract the voltage offset from the raw signal. As stated in the introduction, the voltage offset basically arises due to leakage currents through the sample. Those leakage currents change slowly with a characteristic time-constant which is usually much more than 500 ms. The shot noise associated with the offset has a much higher frequency. Nevertheless, it may be reduced either by cooling the device, or by applying to the sample an appropriate reverse bias. Furthermore, the SEM parameters (scanning frequency, magnification, sample orientation, etc.) can always be chosen in such a way that the time required by the primary beam for crossing the junction is less than 300 ms. Basing on these experimental observations, the offset component can be easily extracted by computing the average value of the raw signal over a period of 300-500 ms. This operation is realized by the averager (OA3 and OA4) in Fig.4 which has a characteristic time constant of 330 ms. If ultra-low scanning frequencies (0.05 scans per second) are required, the time-constant of the averager can be increased up to several seconds just by replacing the capacitor C1. By working according to these principles, present filter is able to track even small AC components of the signal (lower percent range) within few hundreds of milliseconds. Finally, the differential signal is amplified by a factor of five, and it is driven to the last stage as a bipolar signal.

*2.3 Impedance transformer*

The last amplification stage is intended to drive the input buffer (auxiliary) of the SEM.

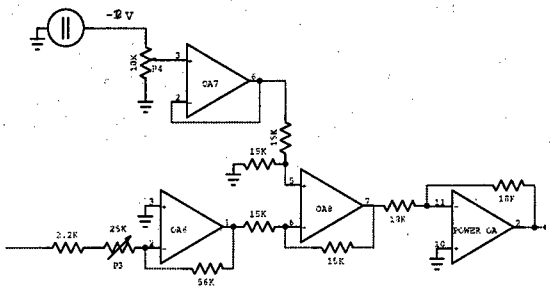



Fig.5 AC Output stage

In the present case a unipolar output with a +5V swing is required which has an output impedance of 75Ω. This function is realized by the differential amplifier AO8 which adds the amplified bipolar signal from the prescaler OA6 with a DC component generated by the voltage follower OA7 (brightness).

## 3. Applications

Several applications of the new EBIC amplifier are shown below and the obtained results are compared with the performances of a dedicated commercial EBIC amplifier with no reverse bias capability. In all cases failed devices are presented, where the (EBIC) signal-to-noise ratio is in the 10%-range.

*3.1 Laser diode*

Fig.6 represents a damaged mirror of a InP laser diode with high reverse leakage current. These pictures have been produced at a very low scanning frequency (typically 0.05 frames per second) with a probe current of 50 picoamperes, an acceleration voltage of 12 kV, and a polarization voltage of 1.5 V. The scanning direction has been choosen in such

a way that the electron beam crosses perpendicularly the active region of the device. It can be clearly seen that the new amplifier resolves the damaged active region (dark central area), being able to image features into the micrometer range. Neither memory effects, nor low-frequency cut-off are observed. On the contrary, the commercial amplifier shows a limited collection capability, by resulting in a very faint EBIC contrast even with probe currents of several nanoamperes.

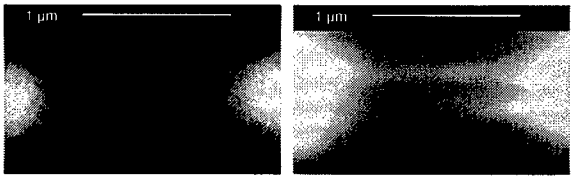



Fig.6 Damaged mirror of a laser diode with the commercial (left) and with the new EBIC detector (right)

### *3.2 Bipolar transistors*

Fig.7, shows a finger of an interdigitated bipolar transistor damaged by an electrostatic discharge and whose emitter-base junction presents a reverse leakage current of 500 microamperes.

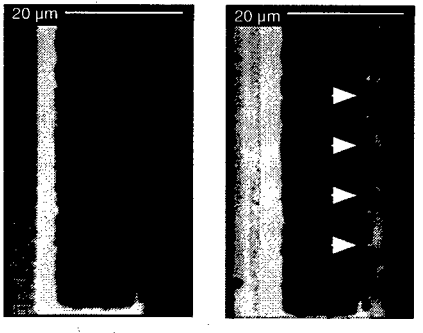



Fig.7 Thermomigration (arrows) in a damaged transistor by the commercial (left) and new amplifier (right)

The junction breakdown site has been previously located by emission microscopy. Since the thermomigrations paths are covered by a thick dielectric layer, the EBIC map has to be produced at fairly high acceleration voltages (typically 18 kV). Under these conditions, the EBIC signal originating from the anomalous recombination sites is superimposed to a very intensive background signal produced by the emitter-base junction, so that a very high contrast capability is required from the collection system. These situation can be easily seen in Fig.7 (left), where the commercial amplifier is deeply saturated by the brilliant background and as a consequence the thermomigration paths cannot be resolved. On the contrary, due to the improved collection efficiency, the new EBIC-detector can be operated at lower acceleration voltages and at lower probe current levels. This property combined with the ability to subtract on-line constant signal levels (self-compensation) improves the signal-to-noise ratio, enabling to image very clearly the anomalous recombination centers.

The emitter-base junction of the transistor in Fig.8 exhibits a low-ohmic short circuit path (due to an electrostatic discharge) which cannot be observed from the top of the device.

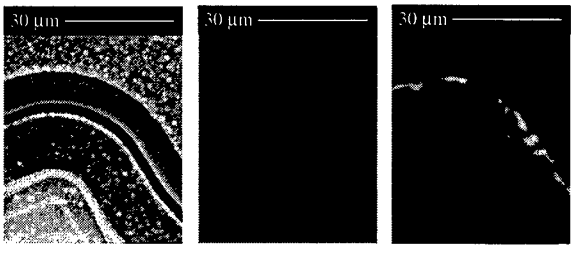



Fig.8 Base-collector junction, topography (left), commercial EBIC amplifier (center), new amplifier (right)

The image resulting by using the commercial amplifier shows a very low contrast even at high acceleration voltages and at high beam currents. The contrast can be improved by applying a low reverse polarization voltage (2.5V) which increases the collection efficiency.

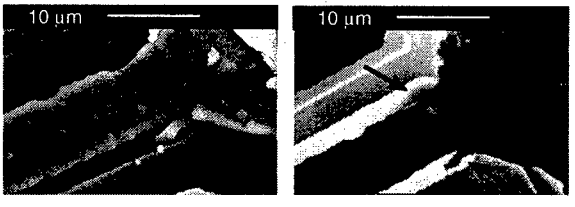



Fig.9 Dark spot and filamentation in bipolar transistor, secondary electrons (left), new EBIC amplifier (right)

It should be noted that in this case the reverse voltage results in a relevant leakage current through the

amplifier (800 microamperes) which can be filtered-out by the adaptive amplifier when the electron beam (13 kV, 2 nanoamperes) is scanned across the sample at a very low scanning rate (typically 0.05 frames per second).

A similar situation can be found in Fig.9 where a dark spot (upper side), due to aluminum thermomigration into silicon, can be observed at the same time with a silicon filamentation (lower side).

### *3.3 Light emitting diode*

EBIC mapping of GaAs light emitting diodes (LED) in planar technology is usually very difficult since before reaching the active region, the primary electron beam has to go through the contact layer and the cladding layer. Therefore, acceleration voltages in the 40 kV range and beam currents (typically 20 nanoamperes) are required. Such a case is shown in Fig.10, where a LED with a reduced quantum efficiency is represented. A photoluminescence map of the device reveals that the LED is emitting in the upper right corner, only (Fig.10 center), while the residual area is undergoing non-radiative recombination. Furthermore, Fig.10 (right image) shows how the roughness of the contact-layer results into a grainy EBIC image. In this case, the significant EBIC signal is the (spatial) low-frequency component of the map. Thus, such map is only possible by means of an amplifier with high sensitivity, high contrast capability, and with a very low- cut-off frequency.

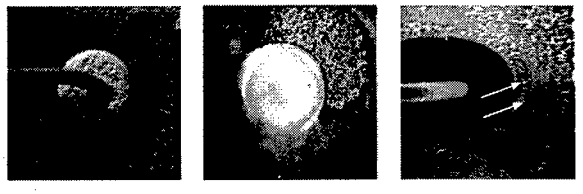


Fig.10 Dark lines and spots in a LED, topography (left), photoluminescence (center), EBIC map (right)

Fig.10 demonstrates that the proposed self-compensating amplifier basically behaves like a DC-coupled EBIC amplifier, which enables to identify dark spots and the dark lines which are located at the border between radiating and non-radiating regions. At the same time the self-compensation avoids that the amplifier goes into saturation if bright areas are encountered. Finally, the EBIC map shows an excellent match with the photoluminescence map.

## 4. Summary and conclusions

A new self-compensating EBIC amplifier has been proposed which enables sample polarization for improved carrier collection and detector's sensitivity. Although the presented detector has the structure of DC-coupled amplifier, it actually combines the properties of an AC-coupled amplifier and especially a very good immunity to leakage currents through the sample.

The performances of the new amplifier have been compared with those of a commercial EBIC detector with no polarization capability and have been tested in the case of critical samples.

## Acknowledgments

The authors are indebted to Mr. A. Orzati for the helpful discussions.

## References

[1] J.F. Bresse
*EBIC in silicon planar p-n junctions: physical model of carrier generation*
Scanning Electron Microscopy Symp (1972)105

[2] B.G. Yacobi and B.M. Ditcheck
*EBIC characterization of multiple in situ three-dimensional junctions in silicon*
Microscopy of semicon. mat. (1987)703

[3] A.E. Luk'yanov and D.V. Kolmikov
*A simple EBIC-signal amplifier with biased sample*
Scanning 14(1992)231

[4] F.J. Maher and C.J. Rossouw
*Design and performance of an amplifier for EBIC imaging in a scanning electron microscope*
J. of Phys. El. Sci. Instruments E16(1983)1238

[5] H.J. Leamy
*Charge collection scanning electron microscopy*
J. of Applied Physics **53**(1982)R51

[6] S.K. Milhstein, D.C. Joy, S.D. Ferris, and L.C. Kimerling
*Defect characterization using SEM-CCM*
Physica Status Solidi A **84**(1984)363



PERGAMON

Microelectronics Reliability 38 (1998) 895–899

MICROELECTRONICS RELIABILITY

# Detailed Investigation of SEM-results by TEM at one Sample Using FIB-technique

**U. Mühle, A. Wiesner, S. Schray**

*Siemens Microelectronics Center Dresden, Königsbrücker Straße 180, 01099 Dresden*

## Abstract

We show how different analytical methods were applied to one sample to get maximum information. By SEM strings of an unknown material in the centre of the array, where bitline shorts appeared, were detected. For the identification of the material and the technology step which caused the fail the modification from the compact SEM-sample to a thin foil for TEM-analysis was necessary. The most critical steps were executed by FIB. Using analytical TEM the origins of the shorts were determined as different materials.

© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

High yield as an indicator for an efficient production of memory chips demands a highly developed failure analysis to clear up the physical reason of electrical fails. One important type of fail are shorts between adjacent metal lines, each responsible for the breakdown of a number of single cells.

An electrical short may have various physical reasons corresponding to different technological process steps. A detailed physical failure analysis of the detected electrical fails should give an indication, how to avoid them. Our examples illustrate, how shorts between two metal lines of a memory chip, showing the same or very similar electrical properties, may have very different physical reasons.

Only the application of different analytical methods to single samples in order to provide maximum analytical information were able to distinct the different technological mechanisms of the fails. Result of the electrical characterization of the breakdown of an array was the number of the failed bitline, but not the location of the fail in the array. A special feature of the electrical results in

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00147-4

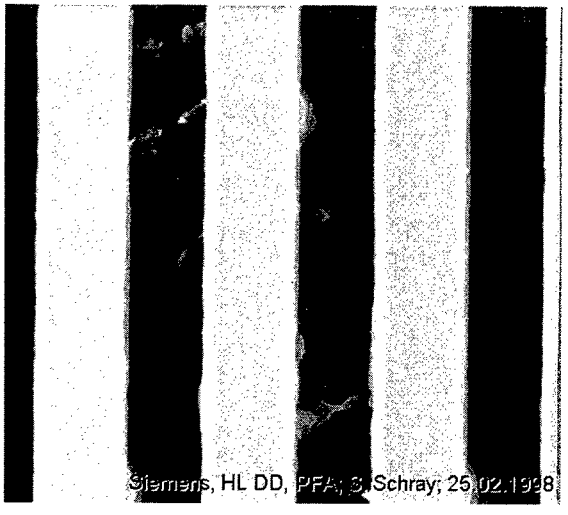



Figure 1: SEM-image of stringers below metal lines, connecting them and leading to a short.

one of our samples was the fact, that the short vanished after application of a voltage, which was a small amount higher than in normal application.

## 2. SEM-investigations

For SEM-analysis it was necessary to remove all layers situated above the metal-one-bitlines to release the fail. This was realized by a combination of chemical and plasma etching steps.

A search along the failing bitline by SEM top view in a Philips XL50-tool lead to the detection of small strings of an unknown foreign material in the centre of the array as the most probable reason for the short. The material was embedded in the oxide surrounding the bitlines (figure 1).

In another case inhomogenities of larger dimensions were found. They can short two

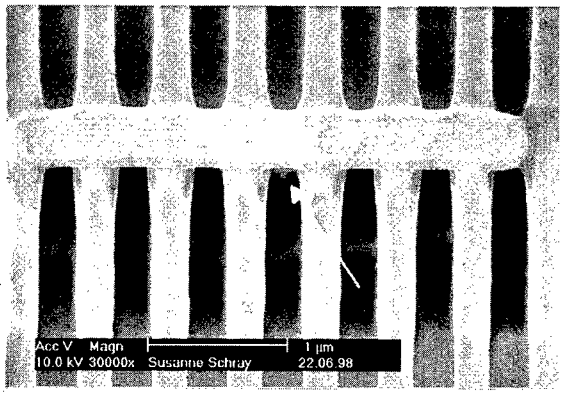



Figure 2: Spot of an unknown material below metal lines (arrow), protected by a FIB-deposited oxide layer

bitlines, but the width can exceed the location of the bitline, as we can see in SEM-image (figure 2).

The material as well as the exact technology step, where the defects were created, had to be identified. The appopriate way for this task should be a TEM-analysis of the discovered fails employing EDX.

## 3. Modification of the samples for TEM

The frequency of this type of fail on the wafer was very low. For this reason it was necessary to use the same sample for the TEM-investigation, which had been used for SEM before. This required a modification of the specimen from the compact one for the SEM-analysis to an electron transmittable foil, which is requiered for TEM. Due to the small dimensions of the defect focussed ion beam technique (FIB) had been taking for marking and for target preparation [1].

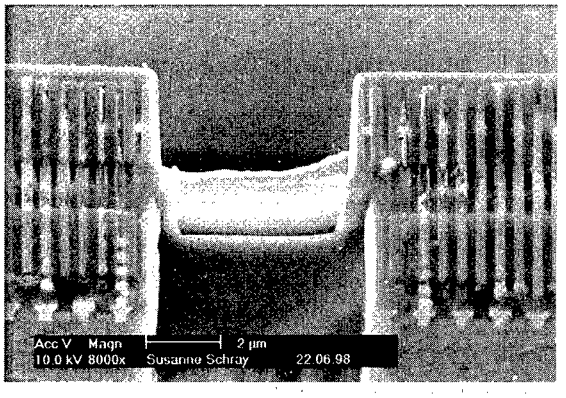



Figure 3: FIB-prepared TEM-lamella, containing the fail

A FEI-tool FIB 800 was used for this purpose.

A very critical step of the preparation in the case of a completely delayered sample is marking, as it is only possible by using ion beam imaging without any protecting layer on the fail. In our application using a small aperture in the FIB allowed to image the location of the fail, so that it could be protected by deposition of platinium or oxide (figure 2). The following step was a mechanical pre-preparation of the marked sample by sawing and grinding.

Final preparation was executed in the FIB again. The result of this preparation is a TEM-lamella, surrounded by compact material (figure3). FIB-prepared TEM-samples have a very high stability, but only a small range, transmittable for electrons.

## 4. Final observation by analytical TEM

For the TEM-investigation a Philips CM 200 with an energy dispersive X-ray-detector by EDAX

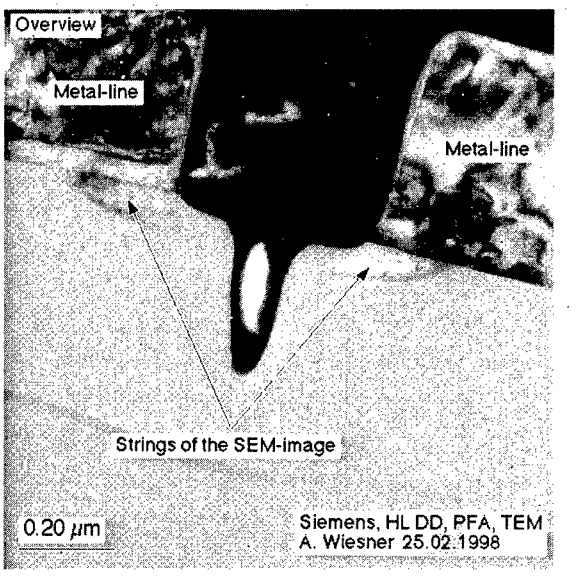



Figure 4: TEM-cross-section of the strings of figure 1. The measured area of the EDX-spectra (figure 5) is the end of the left arrow.

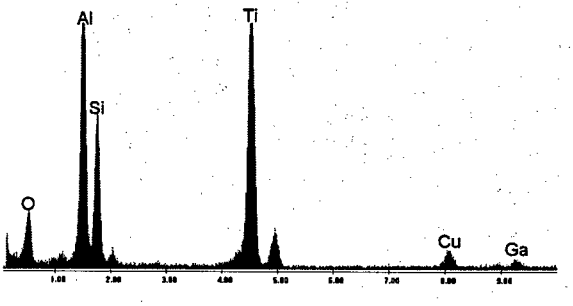



Figure 5: EDX-spectra of the material of the strings (compare to fig. 4). The strong occurence of Ti refers to the origin as liner material.

was used. In the TEM the strings were found again in cross section in the oxide below the two shorted metal lines (figure 4). The bottom of one metal line showed a very low density material with a lot of holes, probably a result of corrosion. This again may explain the absence of the short after

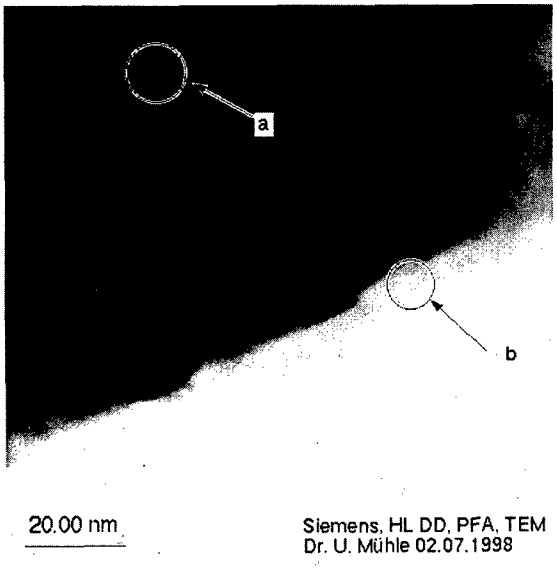



Figure 6: Larger inhomogenity in the oxide below the metal lines, filled by tungsten.

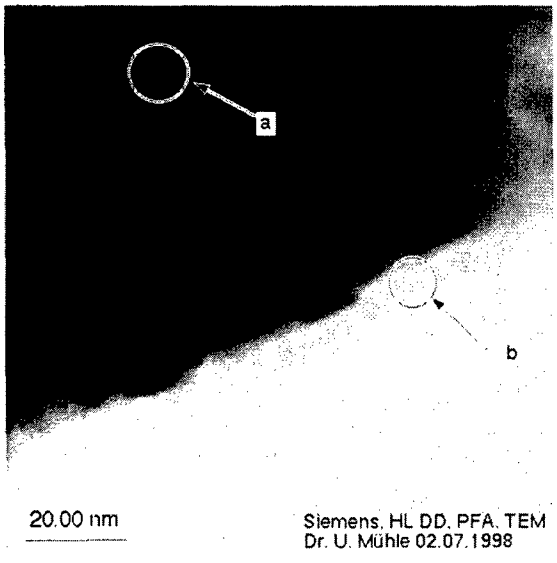



Figure 7: Close up view of the bottom of the fail in figure 6. EDX-measurements at a and b made clear, that the defect in oxide was filled like a regular contact hole with Ti-liner and tungsten.

application of higher voltage, due to a destruction of the electrical contact.

EDX-analysis of the strings provided the origin of the bitline shorting material as a part of the liner deposition material (figure 5). A scratch in the oxide under the metal-one-lines may have been filled during the deposition of the titanium-liner and may thus have connected the two bitlines.

The reason of the fail in the second sample (compare to figure 2) was a larger inhomogenity in the oxide below the metal lines (figure 6). In this layer tungsten filled contacts, surrounded by titanium-liners are situated in other regions of the chip than imaged. The anormal hole in the oxide got a liner like the regular contacts and was then filled by tungsten (figure 7).

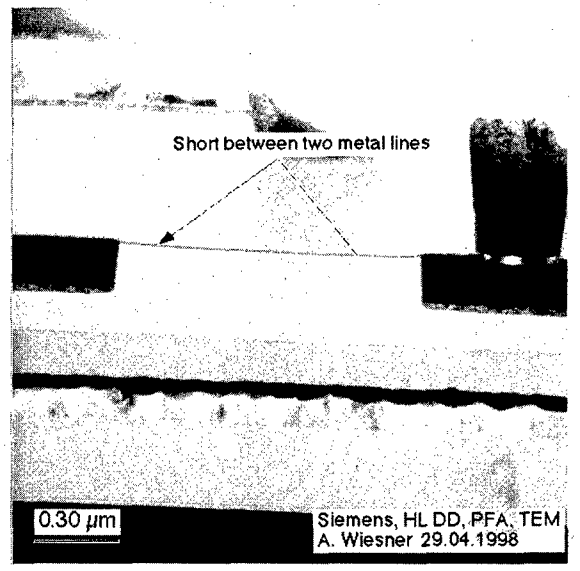



Figure 8: Further example of a short of two metal lines. Very thin residuals of the material of the lines, as identified as tungsten connect them.

In a third case the short was provided by a very thin connection (figure 8), consisting of tungsten (figure 9). The SEM-images of this fail looked very

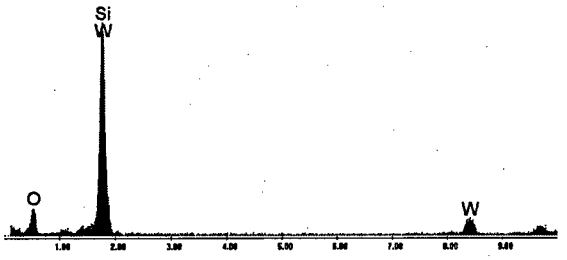



Figure 9: EDX-spectra of the thin connection of the two metal lines in figure 8: The reason of the low intensity of tungsten is the larger dimension of the excited range, compared to the thickness of the conductive connection.

similar to those of figure 2, but they are situated near contacts to higher metal levels.

## 5. Summary

We have seen, that the same electrical fail at memory chips can have its reason in very different physical fails, created in different technological steps. The precise location of the short can be specificated using SEM, but to decide about the origin of the fail it ist necessary to use analytical TEM. In the case of very low density of the fails the further preparation of the SEM-sample to a TEM-foil, containing the spezified detail is the only way for solving the problem.

## Acknowledgment

We want to thank our collegues, providing the electrical investigations. Special thanks to the unnamed contributors for leaving unpublished images.

## References

[1] Fuchs, E.; Oppholzer, H.; Rehme, H.: Particle Beam Microanalysis, Fundamentals, Methods and Applications; VCH Verlagsgesellschaft Weinheim mbH; 1990

PERGAMON

Microelectronics Reliability 38 (1998) 901–905

MICROELECTRONICS RELIABILITY

# Induced damages on CMOS and bipolar integrated structures under focused ion beam irradiation

J. BENBRIK[1], P. PERDU[2], B. BENTEO[1], R. DESPLATS[1]
N. LABAT[3], A. TOUBOUL[3], Y. DANTO[3]

*CNES-SOREP lab, 18 avenue E. BELIN, 33401 Toulouse Cedex 4, FRANCE*
*CNES, 18 avenue E. BELIN, 33401 Toulouse Cedex 4, FRANCE*
*IXL, Université Bordeaux 1, 351 cours de la Libération, 33405 Talence Cedex, France*

**Abstract**

Focused ion beam used for failure analysis or repairing of ASICs degrades device performances by charging up the surface circuit. The influence of focused ion beam on device electrical performances has been studied to understand the behavior of basic test integrated circuits. We have performed measurements on test structures : MOS capacitors and transistors and bipolar transistors. Oxide trap filling and leakage current through the structures induce different failure mechanisms such as : threshold voltage shift in MOS transistors, deep depletion state in MOS capacitors and current gain decrease in bipolar transistors. A characterization of these effects and a first interpretation are proposed. © 1998 Published by Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Focused ion beam (FIB) are widely used for device modification during the debugging phase of the circuit development. Only a few papers are taking into account the fact that the FIB is not harmless for the electronic circuit [1, 2]. Within performing circuit modification with FIB, the use of assisted gas etching or metal deposition creates a local plasma. Degradation of the device electrical parameters is induced by ion charging as it is established in the plasma processing (like RIE, sputtering...) [3]. The surface charge induced by FIB exposure may result in evolution or catastrophic degradation of the integrated circuit (IC) performances.

The purpose of this study was to analyze the focused ion beam interaction with ICs and to find out the main cause of FIB induced damages.

In this paper, we describe the effects of FIB exposure on MOS capacitors and transistors and an evaluation of the electrical parameter shift consequences on the operation of basic cells like inverters. Also, the impact of FIB exposure on bipolar transistor current gain is pointed out. Then, we propose an interpretation of the underlying physical mechanisms responsible for observed degradation.

## 2. Experimental results and discussion

### *2.1. Charging mechanisms*

When the Ga+ ion beam strikes the device, Ga+ ions are implanted only to a shallow depth in the range of 20nm to 50nm. The ion beam interaction with the device surface results in secondary electrons extraction which leaves a positive charge on the passivation layer. The 30keV Ga+ ion beam also builds up a strong positive charge on the passivation layer.

0026-2714/98/$ – see front matter. © 1998 Published by Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00136-X

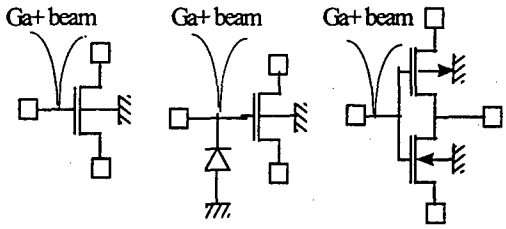



Figure 1: Test circuits

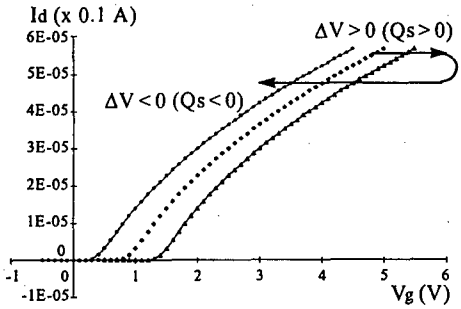



Figure 2 : $V_T$ shift of a floating gate NMOS transistor (Vd = 0.1mV)

Under FIB exposure, the device surface charges up to a high voltage and reaches a saturation value, below 30kV.

Physical phenomena occurring in the device bulk result in charge equilibrium which limits the surface potential. A leakage current establishes between the top of the glassivation layer and underlying conducting layers.

### 2.2. *MOS technology*

MOS devices under test are designed in 0.8µm CMOS technology. The test vehicle includes three structures described in figure 1 : a single transistor, a transistor with a gate protection diode and a basic inverter cell. Let us consider the case of one floating gate transistor. The gate voltage induced by the FIB has been monitored with time exposure using the following method. The experimental conditions are : magnification of 12 with an ion beam current $I_{fib}$ of 50pA and a FIB exposure time of 10 seconds.

The Id-Vg characteristics of the MOS transistors under test are measured as an initial reference before FIB exposure and recorded in-situ during FIB irradiation. Floating gate transistors present a positive threshold voltage shift $\Delta V_T$ observed within several seconds, then $\Delta V_T$ drifts from a positive value towards a negative one which remains permanent (figure 2). MOS transistors which gate is grounded through a P-N junction do not exhibit any parameter shift.

### 2.3. *Estimation of the surface potential*

After FIB irradiation and using the initial characteristics as a reference, we determine the value of the gate voltage required to obtain the final value of the drain current. We find an equivalent gate voltage to be around 20V. We should notice that this experimental parameter is lower than the breakdown voltage which is 37V for this technology.

A simplified equivalent circuit reflecting electrostatic charging mechanisms is used to estimate the floating gate voltage (figure 3).
The presence of a high electric field E in the gate oxide layer is assumed to imply a Fowler Nordheim current crossing the gate oxide, expressed as :

$$J_{fnord} = A\,E^2 \exp\left(-\frac{B}{E}\right) \qquad (1)$$

with : $A = \frac{q^2}{8\pi h}\frac{1}{\phi_0}$, $B = \frac{4}{3}\frac{(2m^*q)^{1/2}}{\hbar}(\phi_0)^{3/2}$

and $\phi_0 = 3.1$eV for electrons, $\phi_0 = 3.7$eV for holes.

Under this assumption, the gate oxide equivalent resistance $R_{ox}$ depends exponentially on Vox voltage (with $V_{ox} = V_g - V_\Phi$) :

$$R_{ox} = \frac{1}{AV_{ox}} \exp\left(\frac{B}{V_{ox}}\right) \qquad (2)$$

Therefore, the transient evolution of charging mechanisms related to the floating gate oxide is defined by :

$$\frac{dV_{ox}(t)}{dt} + \frac{V_{ox}(t)}{R_{ox}C_{ox}} = \frac{J_{fib}}{C_{ox}} \qquad (3)$$

A numerical solution of equation (3) gives an estimation of the surface potential around 500V, and of the floating gate potential of 20V, reached within 1 to 2 seconds and in good agreement with the experimental values.

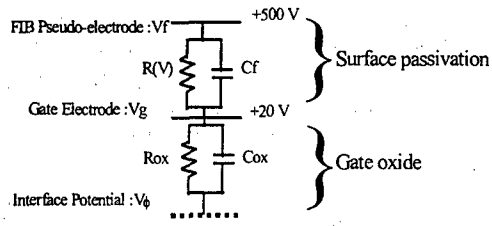



Figure 3 : Equivalent circuit of a passivated floating gate

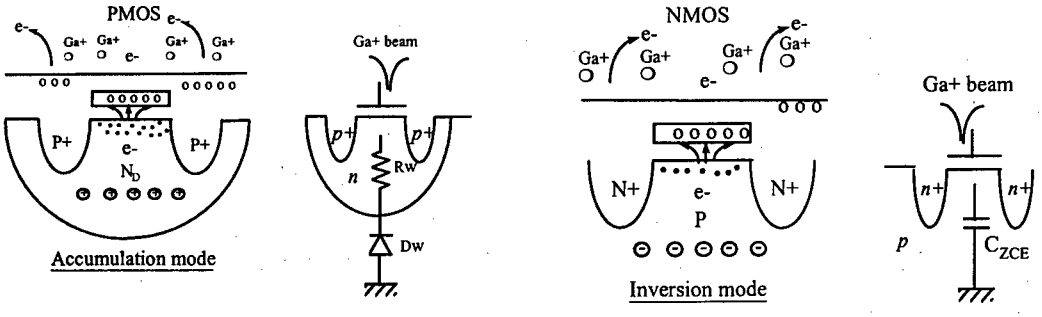



Figure 4 : Charging mechanisms and equivalent circuit of PMOS and NMOS transistors under FIB exposure

### 2.4. *Transient evolution of MOS transistor threshold voltage*

After FIB irradiation, the PMOS transistor remains in an accumulation mode while the NMOS transistor remains in an inversion mode (figure 4). In both cases, electrons are concentrated at the $Si/SiO_2$ interface.

The equivalent circuit of the PMOS transistor comprises a resistor $R_W$ and a diode $D_W$ and for the NMOS transistor a capacitor $C_d$. At the $Si/SiO_2$ interface, $V_\Phi$ is equal to zero for the PMOS transistor and is positive for the NMOS transistor. It results that the gate oxide electric field E is higher in a PMOS transistor than in a NMOS transistor.

Within the first FIB scanning sequences, electrons building up a charge $Q_e$ are trapped at the $Si/SiO_2$ interface while holes ($Q_t$ charge) are trapped close to the Polysilicon gate/$SiO_2$ interface ([4], [5]). The added image charge $\Delta Q_S$, mainly caused by electrons located in the vicinity of the $Si/SiO_2$ interface, is positive. As $\Delta V_T$ is related to $\Delta Q_S/C_{ox}$, the threshold voltage $V_T$ shifts positively. After the initial gate oxide charging, holes trapping contribution increases with FIB exposure time and becomes preeminent. At last, a net positive charge is induced in the gate. It implies the image charge $Q_S$ to become negative and $V_T$ to shift negatively.

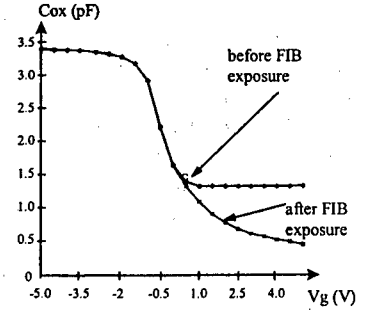



Figure 5 : MOS capacitance vs. gate voltage measured at 1MHz before and after FIB irradiation

### 2.5. *CMOS inverter circuit*

Concerning the CMOS inverter circuit under FIB irradiation., we have observed two cases depending on the location of the leakage current triggering either in the gate oxide of the PMOS or of the NMOS transistor.

If the PMOS transistor exhibits a positive $\Delta V_T$, the NMOS $\Delta V_T$ is negative and reciprocally. Indeed, PMOS and NMOS gates are connected together and submitted to the same voltage. When one transistor presents a gate leakage current and behaves like a resistor, the floating input voltage of the inverter is then maintained to the potential variation across an equivalent leakage resistance. The threshold voltage of the first degraded transistor shifts positively while the other one presents a negative variation of $V_T$.

After irradiation, we had previously verified that PMOS transistors remain in an accumulation mode. The observed sensitivity of PMOS transistor is therefore enhanced by a higher gate oxide electric field than for NMOS transistors.

### 2.6. *MOS capacitors*

Capacitor measurement versus gate voltage was performed with a HP 4280 at 1MHz on large area MOS capacitors, typically 50µmx50µm.

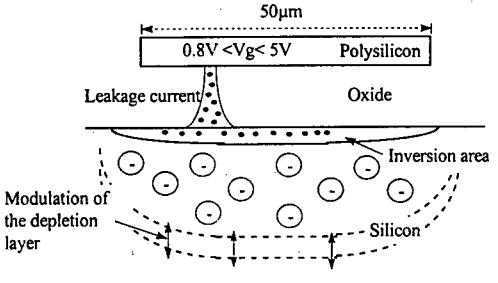



Figure 6 : Capacitor behavior during FIB exposure

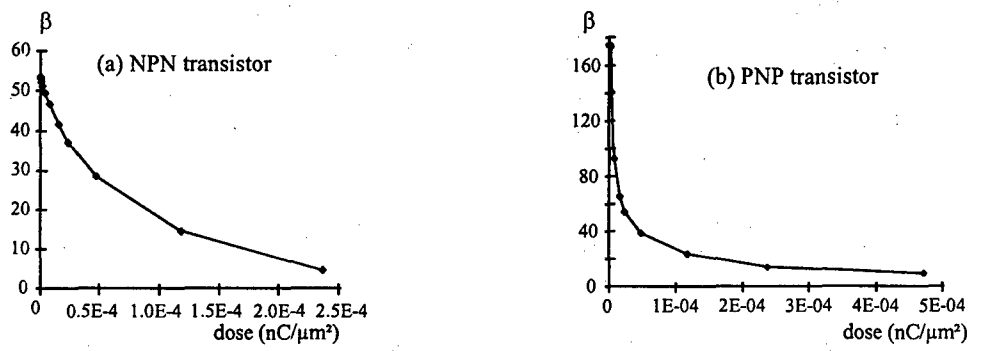



Figure 7 : Decrease of the current gain β with the cumulative dose for NPN (a) and PNP(b) transistors

Usual behavior of the MOS capacitor is observed in the inversion regime before irradiation : the capacitance value does not depend on the gate voltage for 0.8V<Vg<5V. After FIB exposure, a further decrease of the capacitance value with respect to the gate voltage is measured (figure 5).

This behavior, typical from a deep depletion mode, shows that the inversion layer has vanished. In fact, the gate oxide leakage current should evacuate electrons from the inversion region (figure 6). A rise of the gate potential implies an increase of the depletion area thickness and a decrease of the capacitance value.

### 2.7. *Bipolar transistors*

Both NPN and PNP transistors have been exposed under FIB. Experimental conditions are : ion beam current, 50pA, exposed area, 40x40μm$^2$.

Observed current gain decrease are plotted in figure 7. The current gain degradation is attributed to the deformation of p-n junction depletion regions (figure 8) under the passivation layer due to the effect of FIB induced surface voltage.

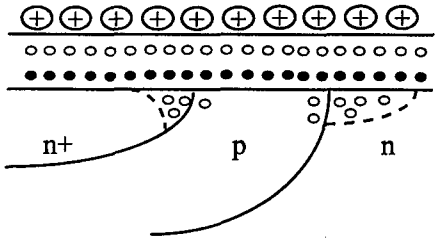



Figure 8 : Effect of FIB-induced charging on depletion regions of emitter-base and collector-base junctions for a NPN transistor

The breakdown voltage increase $BV_j$ of p+-n diode indicates that the junction deserted region is extending (figure 9). The n+-p junction bending leads to an increase of its mean curvature and involves a reduction of the breakdown voltage $BV_j$.

The degradation of the surface region near emitter-base and base-collector junctions can be represented by two parallel diodes under forward bias in the equivalent circuit model of the transistor. This assumption is correlated with the decrease of the bipolar transistor current gain observed during FIB exposure.

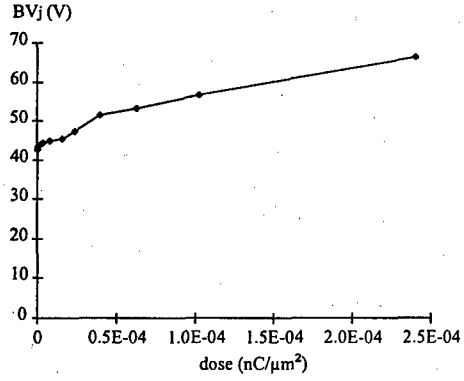



Figure 9 : Evolution of the p+-n junction breakdown voltage BVj vs. under FIB exposure

## 3. Conclusions

FIB induced charging effects induce different degrees of damage in MOS and bipolar devices. The importance of defining safe experimental conditions to perform circuit modification or repairing is pointed out. This work has demonstrated that degradation of MOS and bipolar transistor electrical parameters under FIB irradiation implies capacitive coupling between the device surface and underlying conductors.

Hence, it appears from experiments on MOS transistors and inverters that :

- Some intrinsic mechanisms limit the passivation surface potential.

- A threshold voltage shift is induced by modulation and/or creation of interface traps and carriers trapped in the oxide gate [6].
- The kinetic of the threshold voltage drift during FIB exposure is assumed to result from the establishment of a Fowler-Nordheim leakage current through the gate oxide.
- The FIB irradiation susceptibility of PMOS transistors with a floating gate is higher than for NMOS transistor with a floating gate.

The extent of damage observed in MOS transistors depends on the connection mode of the gate. A gate connected to a diode is sufficient to protect transistors against FIB degradation. This could suggest a robust protection for digital circuits.

## References

[1] R.G. Lee and J.C Morgan, "Integration of Focused Ion Beam System in a Failure Analysis Environment", Proc. of the 17th Int. Symp. for Testing and Failure Analysis, ISTFA'91, Nov. 1991, pp 85-95

[2] Ann N. Campbell, Kenneth A. Peterson, Daniel M. Fleetwood, Jerry M; Soden, "Effect of Focused Ion Beam Irradiation on MOS Transistors", Proc. 35th Annual International Reliability Physics Symposium, IRPS'97, April 1997, pp 72-81

[3] K.G. Druijf, J.J.M. de Nijs, E. Van der Drift, "Radiation damage in MOS structures due to plasma processing", Microelectronics Engineering 22, 1993, pp 231-234

[4] M.V. Fischetti, Z.E Weinberg J.A. Callise, "The effect of Gate Metal and $SiO_2$ Thickness on the Generation of Donors States at the Si-$SiO_2$ interface", J. Appl. Phys. 57 (2), 1985, pp 418-425

[5] Ann N. Campbell, J.M. Soden, J. Rife, "Electrical Biasing and Voltage Contrast Imaging in a Focused Ion Beam System", Proc. 21th Int. Symp. for Testing and Failure Analysis, ISTFA'95, Nov. 1995, pp 33-41

[6] D. J. DiMaria, E. Cartier, D. A. Buchanan « Anode hole injection and trapping in silicon dioxide », J. Appl. Phys. 80 (1), 1996

PERGAMON

Microelectronics Reliability 38 (1998) 907–912

MICROELECTRONICS RELIABILITY

# Junction delineation and EBIC on FIB cross section

G. Perez [a], F. Courtade [a], B. Benteo [b], J-L. Gauffier [c], JL. Kwang [d]

[a] *CNES, Electronic Analysis department, 18 avenue E. Belin - 31 043 TOULOUSE, France*
[b] *SOREP, BP 5, ZI de bellevue, 35 220 CHATEAUBOURG, France*
[c] *Department of physics INSA, complexe scientifique de Rangueil, 31 077 TOULOUSE, France*
[d] *MOTOROLA, BP 1029, Avenue Général Eisenhower, 31 023 TOULOUSE, France*

**Abstract**

This paper presents a sample preparation study on a FIB based X-section. It shows that wet etching is a simple and reproducible solution for the decoration of both implanted and deposited layers. The methodology is applicable on bare or packaged dice.

Results with EBIC on FIB cross-sections were also obtained. A specific double FIB box method was developed and characterized to improve spatial resolution of EBIC. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Deposited layers and junction delineation

The use of FIB for technological analysis of semiconductor is not new. FIB sections are the most widely used application in technology analysis today. This technique was introduced by K. NIKAWA on a SEIKO commercial machine in 1989 [1]. It allows preparation and observation of vertical sections on integrated devices, and opened up a range of application outside those which can be obtained by traditional approach : cleavage and mechanical lapping.

However, the low contrast between the layers in FIB cross-sections can make the feature of interest difficult to observe, which has become a limitation for FIB cross-sectionning.

Historically, failure analysts have tried various methods for layer decoration inside a FIB box :

- GAE (Gas Assisted Etching)
- Dry etching
- Wet etching

*Gas assisted Etching*

GAE is conducted directly inside the FIB chamber. MICRION firstly introduced this method with XeF2 gas [2], it is applicable to deposited layers delineation. Some reproducibility difficulties have been emphasized by some authors with this methodology [3].

*Dry etching*

Dry etching (plasma mode) gives good results for deposited layers but poor ones for junction delineation (AMD) [4]. This sample preparation cannot be used with packaging and is therefore a severe limitation for our applications which are at customer level (finished products).

*Wet etching*

Wet etching on FIB cross-section has sometimes been mentioned (CNET-CEA) [5] but to our knowledge no attempt has been made to optimize this specific etching effect.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00104-8

The first goal of this paper is to achieve the best control of the decoration process for both deposited layers and junction delineation. The methodology has to be applicable on bare or packaged dice.

### *1.1. Decoration procedure*

In a FIB X-section decoration there are two main difficulties :

1) The FIB box
2) The $Ga^+$ implanted layer

*1) The FIB box*

A conventionnal FIB box is a closed area. It is really different compared to the classical cross-sectionning technique : mechanical lapping or cleaving where a large surface is exposed during the chemical dipping.

Due to the size and depth of the box, wettability problems can be encountered and repercussions on the quality of the decoration may be significant.

*2) The $Ga^+$ implanted layer*

Inside the FIB box, $Ga^+$ ions are implanted. In order to choose beam parameters we have characterized this $Ga^+$ layer. EDX results show the variation of $Ga^+$ implanted, versus beam current (figure 1.a & 1.b).

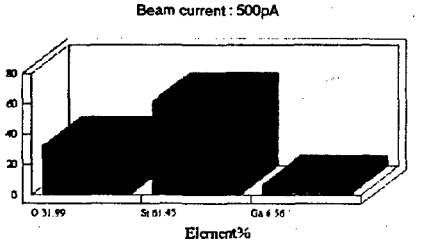



Fig 1.a

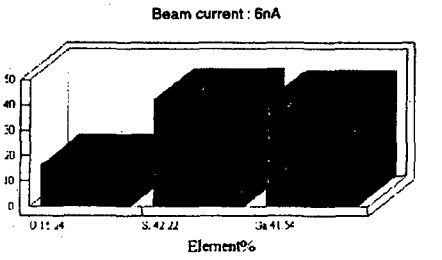



Fig 1.b : $Ga^+$ versus beam current (experimental results on a silicon sample)

At the beginning of the study our first approach was to remove the implanted $Ga^+$ layer before any delineation test. For that we tried common industrial solution which is used for mask repair [6].

In the field of FIB mask repair the photolithography specialist came across the problem of $Ga^+$ implant suppression. In this process ions are used to sputter mask defects which are in the form of chromium thin film protuberances.

During sputtering process ions are implanted into the quartz substrate and in so doing locally affect the transmittance of the mask which becomes unusable.

In ionic lithography, chemical reactives have also been used for the etching of the amorphous layer induced by ion implantation [7]. Unfortunately these methodologies are not simple to implement in failure analysis sample preparation processes.

The major difficulty is the lack of selectivity of etchants like hot $H_3P0_4$ with the different layers of the X-section.

### *1.2. Test sample*

Our second approach is to develop a test sample in order to try a kind of dichotomy of the problems mentioned above related to the FIB box.

This test sample (figure 2) is fitted with :

- A mechanical lapping section
- Open FIB box
- Classical FIB box (closed)

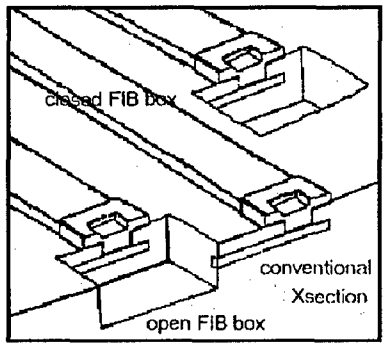



Fig 2 : Test sample

This triple configuration allows us to analyse different effects of wettability and $Ga^+$ amorphous layers. The lapping section gives an element of comparison for reproductibility aspects linked to chemical reactives.

The open FIB box reduces wettability effects, only $Ga^+$ layer can disturb the delineation process. The classical FIB box combines all the above effects.

### *1.3. Experimental results*

Experiments were conducted on test samples in order to determine decoration conditions. Several factors were considered such as :

- FIB box geometry
- FIB current
- Chemical solutions : buffer etch, addition of surfactants to BHF, $HNO_3$/HF 20/1, BHF/$H_2O_2$ which allow to achieve high wettability to Si surfaces without adding any surfactants
- Decoration time

In figure 3, it can be noted how usefull the test sample is. In this particular case, junction inside the open box is not fully delineated (A), compared to the result at the mechanical lapped surface (B).

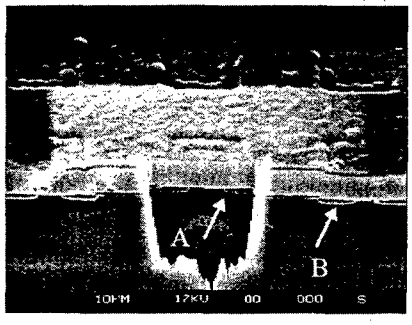



Fig 3 : Open FIB box and mechanical lapped surface

In figure 4, another example shows delineation of junction and decoration of layers fully achieved in a closed FIB box configuration.

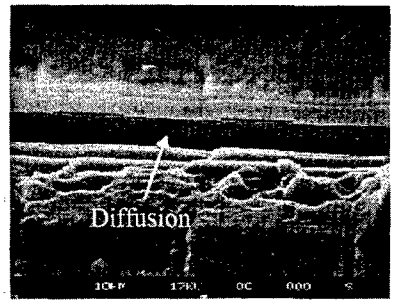



Fig 4 : Detail of closed FIB box inside the test sample

Best results were obtained by using the well known HF/$HN0_3$ 20/1 mixture.

For junction delineation dipping duration must be shorter than with lapping sections. A dip as short as possible is the most efficient ( $\approx$ 1s). This may be due to the $Ga^+$ implanted on the semiconductor which enhanced the decoration process. This phenomenom has already been cited in literature for specific ion beam lithography applications. This enhancement effect seemed to be stronger than wetting behaviour due to the small FIB box. For layer decoration the opposite situation has been observed, and dipping duration must be longer (two or three times longer).

### *1.4. Applications*

A typical decoration and observation procedure can be described as follows :

- cross-sectionning using FIB system P2X by SCHLUMBERGER
- removing the sample from the FIB and etching it under pre-defined conditions in 20/1 mixture (see details above).
- transferring the sample into a high resolution SEM (in the paper results presented come from an old S250MK3 by LEICA. We are now using a new LEO982 Gemini).

We applied these recipes to conventional FIB boxes for deposited layers decoration and junction delineation. We succeeded with the deposited layers decoration (figure 5 & 6) and with junction delineation on bipolar, MOS, CCD and hardened technology (figure 7,8 & 9).

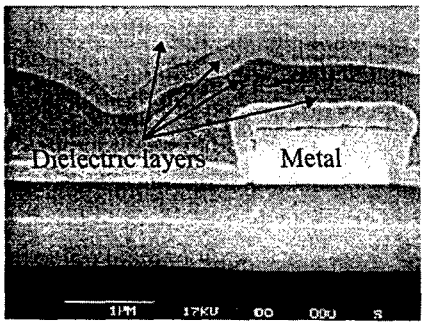



Fig 5 :View of metal and dielectric layer

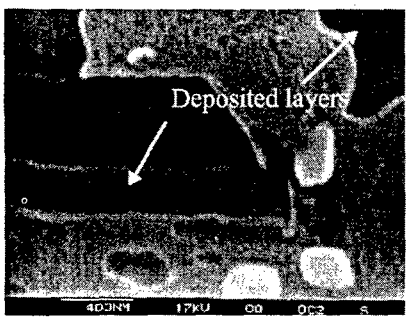



Fig 6 : Detail of deposited layers

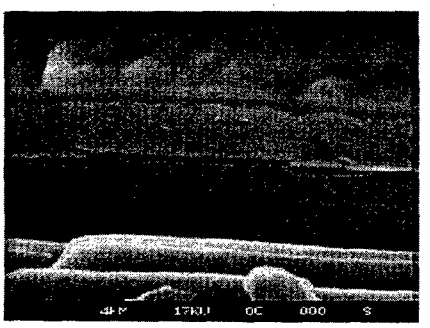


Fig 7 : Detail of the channel area of a TMOS power FET

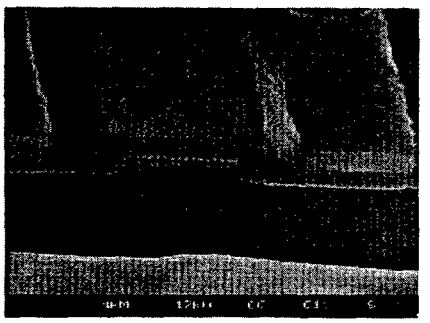


Fig.8 : MOS on a CCD

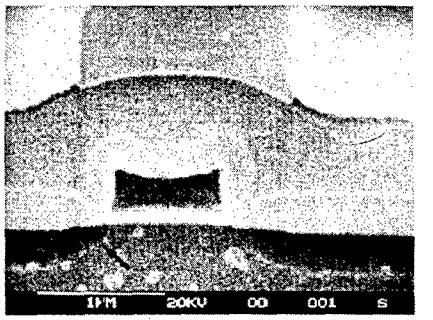


Fig.9 : LDD on MOS transistor

## 2. EBIC on FIB Cross-section

EBIC measurements in failure analysis and device diagnostics fall into two groups. One type of analysis involves locating a defect in a device or junction, or locating an irregularity in a junction.

The other type of EBIC analysis involves measuring depletion widths, junction depths, channel lengths, or in some way utilizing the EBIC signal to characterize the device of interest.

Our goal is to show that EBIC on FIB section can be helpful in the two types of analysis mentionned above.

EBIC analysis on traditionnal cross-section is very difficult due to global deterioration of the sample and damaging of the electrical connexions. For this reason, it has only been applied in a limited number of cases [8]. With FIB section the sample can be prepared properly in order to avoid sample damages and to achieve meaningful results.

### *2.1. EBIC image of FIB sectionned transistors*

EBIC feasability trials on classical FIB cross-sections were carried out on bipolar transistors. Surface EBIC response is not disturbed by the FIB box (figure 10).

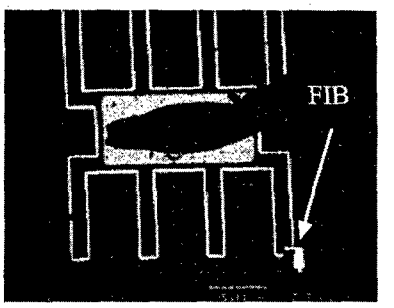
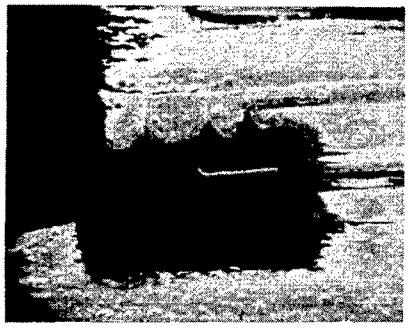



Fig 10 : FIB box doesn't disturb surface EBIC

EBIC on FIB microsection can be obtained, a good resolution can be achieved (figure 11).



2N2369 transistor junctions

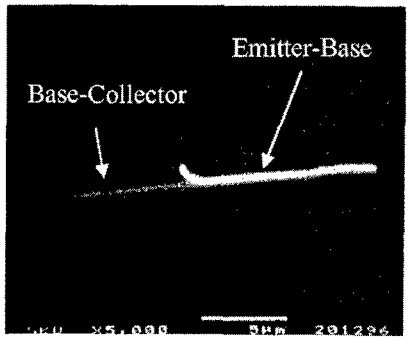



Fig 11 : detail view of 2N2369 transistor junctions

We currently have analysis of mixed ASIC technology and ESD zapped circuits. We hope to observe an in depth EBIC response to complement the classical surface signature.

*2.2. Diffusion length of minority carriers*

In electronic devices, EBIC technique makes it possible to visualize PN junction which are perpendicular to the surface.

In this case diffusion length of minority carriers can be measured on either side of the junction.

The component studied in our example is a NPN bipolar transistor (2N2369).

Two FIB boxes were machined very close together in order to obtain a thin layer of Silicon 1,7µm thick. This double box was placed in the emitter base junction.

The goal of this configuration is to optimize the spatial resolution of the EBIC signal. As far as the authors know the combination of EBIC on this double FIB box configuration has not been done before.

In order to locate the emitter/base junction, we produced SEI and EBIC images, the mixed images allowed us to estimate the position of the junction. Images were obtained with a tilt of 0°, 20° and 35°.(figure 12).

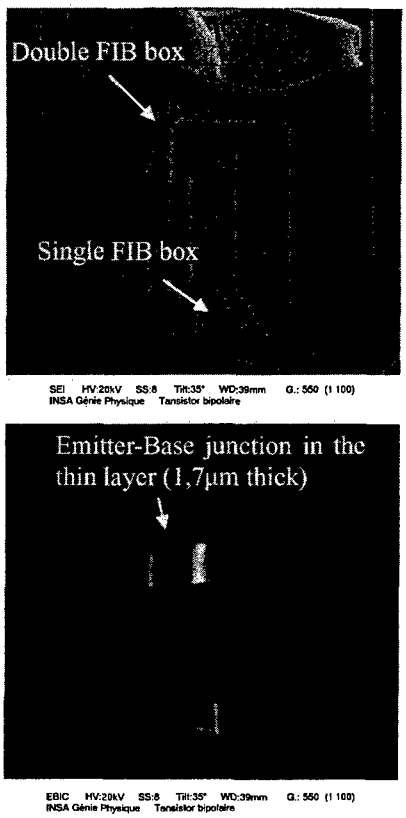



Fig 12 : SEI and EBIC views

The diffusion length of minority carriers in the vicinity of a junction which is perpendicular to the surface was reported by MARTEN and HILDEBRAND in 1983. In our case, relationship between electron beam, excitation volume, depletion region of PN junction is shown figure 13.

Electron/Hole pairs generated by the beam create an EBIC current if they are located at edges of the depleted layer of the junction. EBIC current generation in this region may be written :

P type side : $I=I_0 \exp(-x_n/L_n)$
N type side : $I=I_0 \exp(-x_p/L_p)$
$x_n$ and $x_p$ : distance to the junction

$L_n$ and $L_p$ : carriers diffusion length

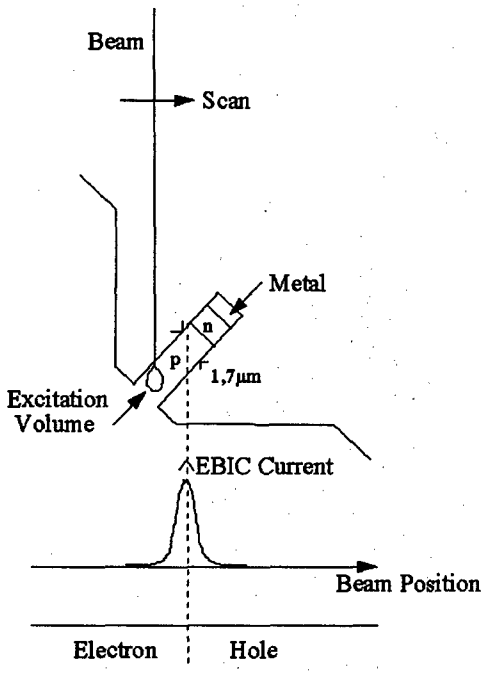



Fig 13 : EBIC configuration inside the double FIB box region

Data analysis shows up the validity of the model, however the diffusion length which where founded $L_p$ = 0,56µm and $L_n$ = 0,83µm are not significant (figure 14).

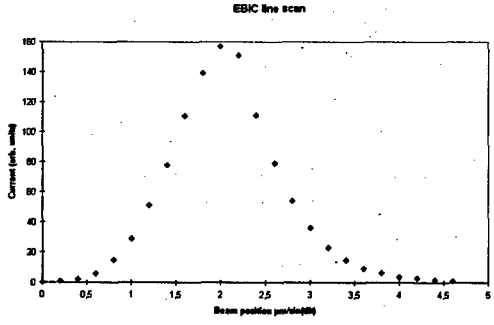



Fig 14 : EBIC line scan

We consider that in the thin active region we studied, carrier diffusion is limited by surface recombinaison caused by the lateral wall.

In conclusion, results obtained show that surface defect produced by FIB (amorphous layer and $Ga^+$ implant) do not affect the EBIC junction localisation process.

Minority carriers diffusion length measurements seem to be in accordance with the geometry of the sample studied.

$L_n$ and $L_p$ measurements are in progress on a single "FIB box" in order to validate these results.

## 3. Conclusion

The wet etch is a simple and reproducible solution for the decoration of both implanted and deposited layers on FIB cross-sections. The methodology is applicable on bare or packaged dice. The feasibility of EBIC on FIB cross-section has been demonstrated. A new double FIB box method for EBIC resolution improvement has been introduced.

## Acknowledgments

The authors would like to thank David MEUNIER and Bernard BARADAT for their helpful comments on the manuscript and word processing.

## References

[1] New application of focused ion beam technique to failure analysis and process monitoring of VLSI
K. NIKAWA, K. NASU and M. MURASE - IRPS 1989
[2] Gas assisted etching with focused ions beam technology - MICRION - ESREF 93
[3] Advanced FIB applications for product analysis
MOTOROLA East Kilbride - SPIE 1994
[4] How to decorate FIB cross sections using plasma etch for SEM observation
SXLI, A. Gray - Advanced Micro Devices - ISTFA 95
[5] Observation of submicronic contacts and via by scanning ions and electron microscopy
R. Pantel, G. Mascarin and J-P. Gonchond - CNET
D.Lafond - CEA - ESREF 93
[6] Advances in focused ion beam repair of opaque defects
IBM - SPIE 2194
[7] Maskless etching of a nanometer structure by focused ion beams
M.Komuro, H. Hiroshina - JVST B1 (4), 1983
[8] Low acceleration voltage EBIC using FESEM and application to cross-sectional junction evaluation
Y. KADOTA, K. MORI - ISTFA 96

PERGAMON

Microelectronics Reliability 38 (1998) 913–917

MICROELECTRONICS RELIABILITY

# Direct Observation of Local Strain Field for ULSI Devices

N. Hashikawa, K. Fukumoto, T. Kuroi, M. Ikeno and Y. Mashiko

ULSI Development Center, Mitsubishi Electric Corporation

Mizuhara 4-1, Itami, Hyogo 664-8641 JAPAN

**Abstract**

The in-depth profile of strain distribution from the silicon surface is one of the most important pieces of information for optimizing the device performance. The convergent-beam electron diffraction(CBED) method has been applied to analyze the local strain filed of the active regions for both test structure with the shallow trench isolation(STI) and the conventional LOCOS on a cross-sectional surface. As a result, strain distribution was observed successfully. It was found that the compressive stress exist all over the survey regions. The active region close to the bottom corner of the STI shows a larger stress than that of the conventional LOCOS. It is demonstrated that the CBED technique is very effective for the determination of local strain field in a small area of semiconductors and the optimizing of the STI structure and fabrication process.

© 1998 Published by Elsevier Science Ltd. All rights reserved.

## 1. Introduction

In recent years, local strain field in the highly integrated ULSI devices become one of the most significant factor reducing a device performance. A large strain field can influence the electrical and optical properties of the composing materials and it sometimes induce the crystal defects in the silicon substrate[1]. Therefore, it is necessary to evaluate the local strain with spatial resolution nanometer levels for ULSI devices.

For the measurement of the lattice strain, X-ray diffraction and microscopic Raman spectroscopy have been used. However, the spatial resolution of these methods is insufficient for a detailed analysis of ULSI devices. On the other hands, the CBED method has been realized as a extraordinary technique to obtain the crystallographic information with a high spatial resolution due to a small probe size[2,3,4]. In the CBED patterns, higher order Laue zone(HOLZ) lines are observed. Since the positions of HOLZ lines in the CBED patterns depend on lattice parameters, the CBED technique has been used to measure the local lattice parameters. In recent years, Energy-filtered transmission electron microscopes(EF-TEM) combined with the energy filtering system has made it possible to acquire the energy-filtered CBED patterns. The quantitative information can be obtained from energy-filtered CBED patterns because of higher contrast of CBED patterns realized removing the inelastic scattering.

In the present work, the energy-filtered CBED

0026-2714/98/$ – see front matter. © 1998 Published by Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00135-8

technique is applied to the measurement of the local strain field on the nanometer scale, on a cross sectional surface of semiconductor devices. Both test structures with the STI and the conventional LOCOS are examined for the measurements. This paper also describes in-depth distribution of local strain field relating to the isolation structure by the energy-filtered CBED analysis.

## 2. Experimental procedure

### *2.1. Sample preparation*

The two type of test structures were examined in this work. One was the memory cell pattern which was isolated by the conventional LOCOS structure, and the other was isolated by the STI structure. To prepare the cross-sectional TEM sample for the CBED analysis, the test structure wafer was cut along [011] direction. Then it was thinned by mechanical thinning to a 20 $\mu$ m, and finally milled with Ar ion beam accelerated at 3.5kV at 12 degrees. The observed area for both test structure was nearly constant in thickness.

### *2.2. Energy-filtered CBED observation and Computer simulation*

The observations of the conventional LOCOS structure and STI structure were examined with a TOPCON EM-002B microscope combined with a Gatan imaging filter. This system is able to obtain the energy-filtered images and energy-filtered CBED patterns. The energy-filtered CBED analysis was carried out from a [011] cross-sectional direction operated at 95.4kV with a probe size of 10nm.

The local lattice parameters were determined from the fitting between experimental HOLZ(Higher-Order Laue Zone) line positions and simulated ones. The simulated HOLZ lines are based on kinematical theory[5].

## 3. Results

The energy-filtered CBED pattern in Fig. 1(a) is obtained from the silicon substrate. It is clear that the sharp black lines appear in the disk.These black lines

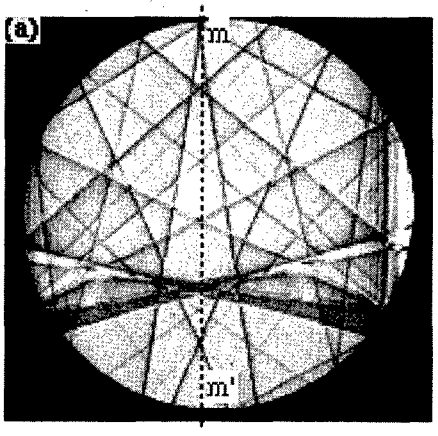



Fig. 1(a). CBED pattern taken from Si substrate. Accelerated voltage is 95.4kV.

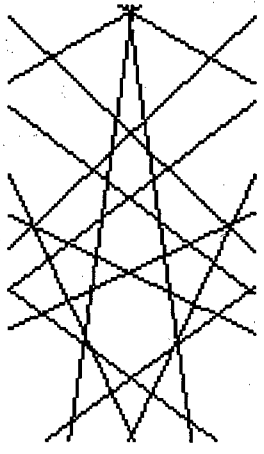


Fig. 1(b). Simulated HOLZ pattern in fig. 1(a) using the kinematical theory.

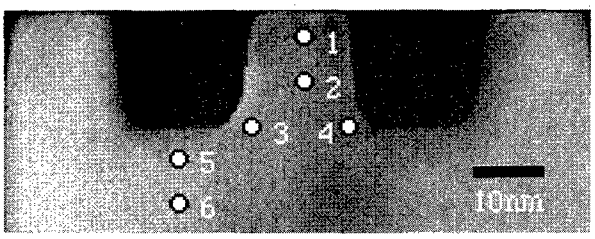



Fig.2(a). Cross-sectional TEM image for STI structure. The energy-filtered CBED patterns are taken from white dots.

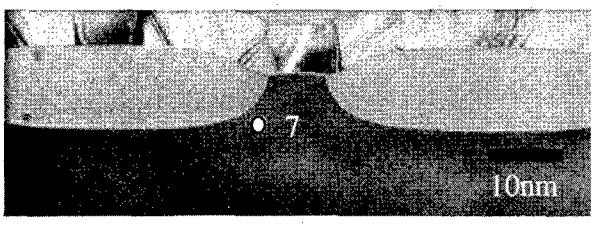



Fig.2(b). Cross-sectional TEM image for LOCOS structure. The energy-filtered CBED pattern is taken from white dot.

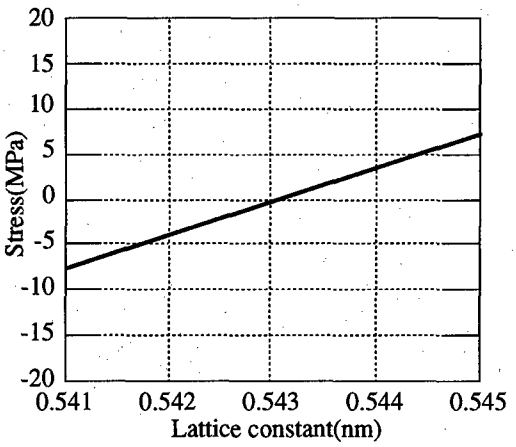



Fig.3. Lattice constant dependence of the stress for silicon.

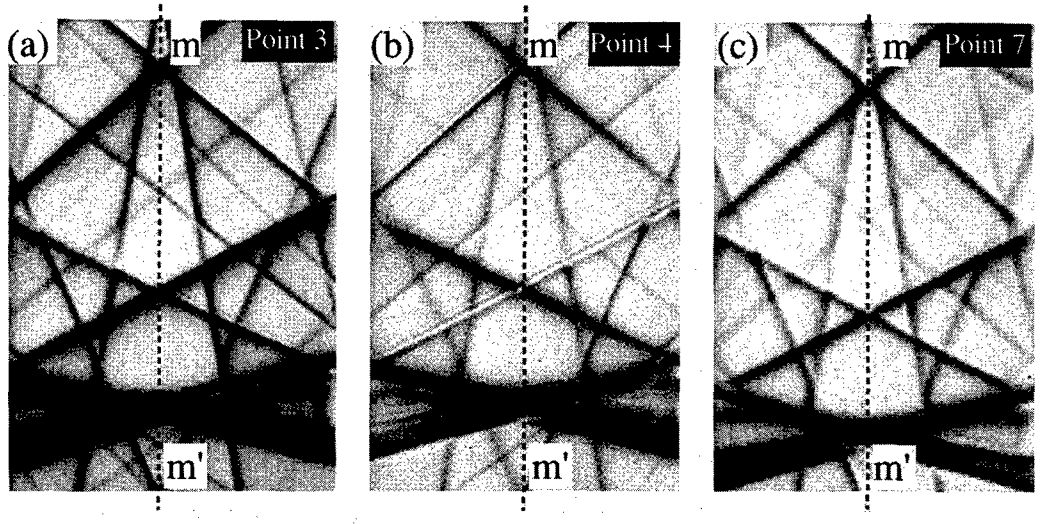



Fig.4 (a) and (b) are the energy-filtered CBED patterns for STI test structure taken from points 3 and 4 in fig.2(a), respectively. (c) is obtained from the point 7 in fig.2(b) for LOCOS structure. The energy-filtered CBED patterns for (a) and (b) are asymmetrical with respect to the line m-m', but (c) displays mirror symmetry.

are the HOLZ lines. The pattern has a mirror symmetry with respect to the line m-m'. The simulated HOLZ lines are displayed in Fig.1(b). The lattice parameters are determined by the comparison of the simulated pattern with the experimental one.

Fig.2 (a) and (b) show the cross-sectional TEM images for the memory cell patterns which are isolated by the STI and the conventional LOCOS, respectively. The white dots indicate the positions where the energy-filtered CBED patterns are obtained.

Fig.3 shows the lattice constant dependence of

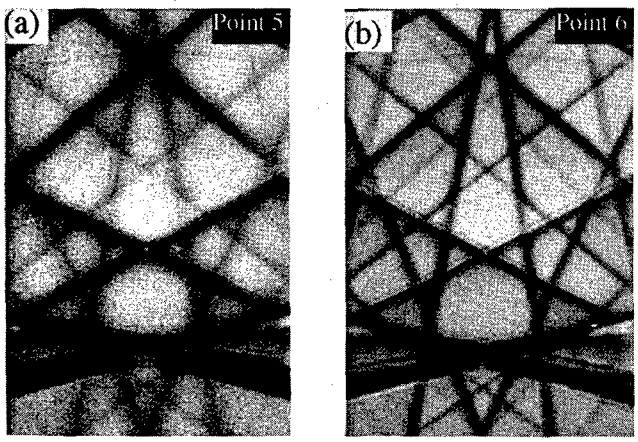



Fig.5(a) and (b) are the energy-filtered CBED patterns obtained from 5 and 6, respectively. Split HOLZ lines can be observed for both patterns.

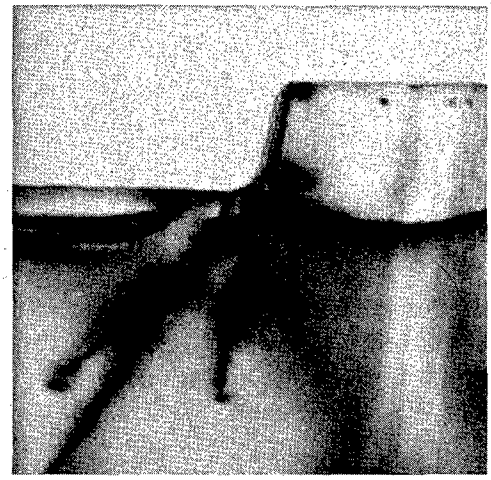


Fig.6. Cross-sectional TEM image for STI test structure. Several dislocation lines can be observed.

the stress for the silicon. The stress values are calculated using the elastic theory[6]. We use this calibration line to evaluate the stress values. The lattice constants of the active region(points 1 and 2) in Fig.2(a) are 0.5415nm for point 1 and 0.5425nm for point2, and the compressive stress values are -5.2MPa for point 1 and -1.4MPa for point 2.

Fig.4 shows the energy-filtered CBED patterns obtained from the active regions close to the bottom corner of the STI(points 3 and 4) in Fig.2(a). In contrast to the above results, the mirror symmetry with respect to line m-m' does not exist. On the other hand, the energy-filtered CBED pattern obtained from the active region close to the bottom corner of the conventional LOCOS structure(point 7) does have a mirror symmetry, as shown in Fig.4(c). This result means that the component of stress is different between the STI structure and the conventional LOCOS structure, and the strain for the STI strcture is larger than the conventional LOCOS structure.

The energy-filtered CBED patterns taken from the bottom of the STI (points 5 and 6 )in Fig.2(a) are shown in Fig.5. It can be seen that these HOLZ lines become split lines due to the strong lattice distortion.

The micrograph of Fig.6 gives the bright-field TEM images for the other memory cell region for STI. The dislocation lines starting from the bottom corner of the STI structure are observed. This result also suggests that the concentrated stress exists close to the bottom corner of the STI.

## 4. Summary

The energy-filtered CBED analysis was the first to apply detection of the local strain field in cross-sectional analysis for the two type of the memory cell regions with the STI and the conventional LOCOS, respectively. The following conclusions are obtained.

(a) The in-depth profiles of strain distribution from the silicon surface is detected by the energy-filtered CBED analysis.

(b) For the STI test structure, the compressive stress exsists in the active region. The active region

close to the bottom corner of the STI shows a larger stress than that of the conventional LOCOS.

It is demonstrated that the local strain field for the ULSI devices can be determined on nanometer scale by the energy-filtered CBED. This method may be a powerful tool for determination of two-dimensional strain profiles, and provide useful information for the optimization the ULSI device.

**References**

[1] Dellith M, Booker G R, Kolbesen B O, Bergholz W and Gelsdorf. J.Electrochem. Soc. 143 pp210-215(1996).

[2] Tanaka M and Terauchi M. Convergent-beam electron diffraction, JOEL, Ltd., Tokyo(1985).

[3] Steeds J and Mansfield J. Convergent beam Electron Diffraction of Alloy Phases, Adam Hilger, Ltd., Bristol, UK(1984).

[4] Maier H J, Keller R R, Renner H, Mughrabi H and Preston A. Philos. Mag. 23 pp74-82 (1996).

[5] Humphreys C J. Rep. Prog. Phys 42 pp1826-1887(1979).

[6] Wortman J J and Evans R A. JAP 36 pp153-156(1965).

PERGAMON

Microelectronics Reliability 38 (1998) 919–924

MICROELECTRONICS RELIABILITY

# A new experimental technique to evaluate the plasma induced damage at wafer level testing

**L. Pantisano[a,b], A. Paccagnella[a,b], L. Pettarin[a], A. Scarpa[a], G. Valentini[c], L. Baldi[c] and S. Alba[c]**

[a] Dipartimento di Elettronica ed Informatica, Università di Padova, via Gradenigo 6/A, 35131 Padova, Italy

[b] Unità INFM di Padova, via Marzolo 8, 35131 Padova, Italy

[c] SGS-Thomson Microelectronics, via C.Olivetti 2, 20041 Agrate Brianza (Milano), Italy

**Abstract**

Plasma processing has become an integral part of the IC fabrication, since it offers advantages in terms of directionality, low temperature and process convenience. However plasma processing induces an oxide charging damage, which is function of process conditions and gate interconnect layout. At the end of the process the plasma damage is "hidden" by hydrogen passivation and becomes latent. At a first electrical inspection all the devices on wafer present nearly the same electrical parameters, whereas a small stress is enough to reveal the plasma damage, producing again a drift of all transistor parameters. In fact the applied stress both depassivate the latent plasma damage and introduce newly defects, which are electrically undistinguishable from plasma damage and depend on applied stress conditions. In our contribution we propose an experimental stress methodology to investigate both the latent damage depassivation effect and the net contribution of plasma damage. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

The use of plasma treatments during the IC manufacturing process induces a stress in the MOS gate oxide. From an electrical point of view, the plasma damage can be roughly emulated by a Constant Current Stress (CCS), which produces oxide charging and $Si/SiO_2$ interface states [1]. Oxide defects lead to a drift of all transistor parameters and affect quality and reliability of the gate oxide (such as Charge to BreakDown ($Q_{BD}$)) [5]. Plasma damage depends on the gate interconnect layout. A charge collecting structure (gate antenna) connected to the gate during the plasma process can substantially enhance the injected charge into the gate oxide [5]. The oxide charging depends on several factors such as the ratio between antenna area and gate oxide area, the antenna geometry, the transistor position on the

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00102-4

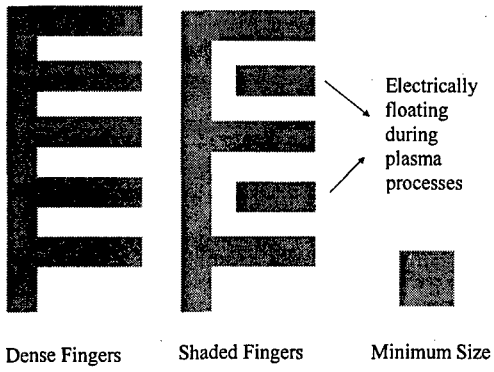



Fig. 1: Gate antenna geometries used throughout this work.

wafer, and the presence of a protection diode connected to the gate. Plasma characteristics play another key role in determining the oxide damage, in connection with the device layout. For instance, Al-etching induced charging strongly depends on antenna perimeter [1], while resist stripping induced damage depends on antenna area [1]. After the Post Metallization Annealing (PMA) step, plasma damage is passivated owing to hydrogen diffusion [2], and all the devices present nearly the same electrical behavior across a wafer. However, even though the PMA can completely hide the oxide damage, a small electrical stress can easily depassivate the latent damage, producing again a drift of transistor parameters [7].

With the purpose of investigating the effect of plasma damage in fully processed structures, it is necessary to reveal the latent oxide damage by using a suitable stress [2,7]. The electrical stress induces both a depassivation of latent damage (plasma damage) and a new oxide damage. For a given stress condition, a major problem concerns the identification of the depassivated plasma contribution within the global stress induced damage [3,5,7].

In this work we have investigated the impact of plasma damage in fully processed structures. We have developed a new experimental method aiming to evaluate the net effects of plasma damage and depassivation through an applied electrical stress The applied stress conditions (Fowler-Nordheim constant current stress) have been chosen so to minimize the stress induced damage and to reduce the stress time.

## 2. Devices and experimental

MOSFETs used throughout this work have been fabricated with a 0.6μm CMOS process with 2 metal levels. The oxide thickness is 12nm, and the device size is W/L=50μm/0.6μm. The gate oxide has been grown in a pyrogenic ambient and subsequently nitridized (nitrided steam oxide). The antennae used throughout this work are shown in fig. 1 and listed in tab. 1. In the shaded M.P.F (fig.1) some teeth of the comb structure are left floating, that is, not connected to the gate metal, so to investigate the shading effect [6]. The dense-fingers antenna have been designed with the purpose of simulating the oxide charging in a dense interconnect layout. In the dense-fingers antenna, both the area and perimeter are twice those of dense fingers ½, and four times those of dense fingers 1/4. All the antennae are present in each test structure designed on the wafer.

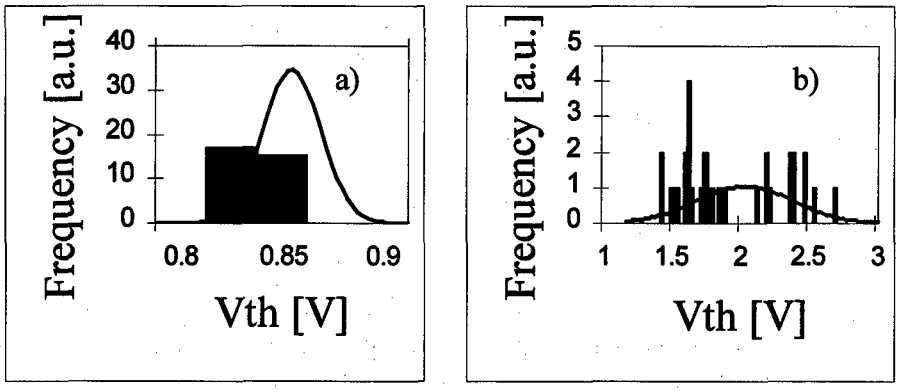



Fig. 2a and 2b: Threshold voltage distribution for a Shaded Antenna before (a) and after stress (b). In solid line are the corresponding gaussion distribution.

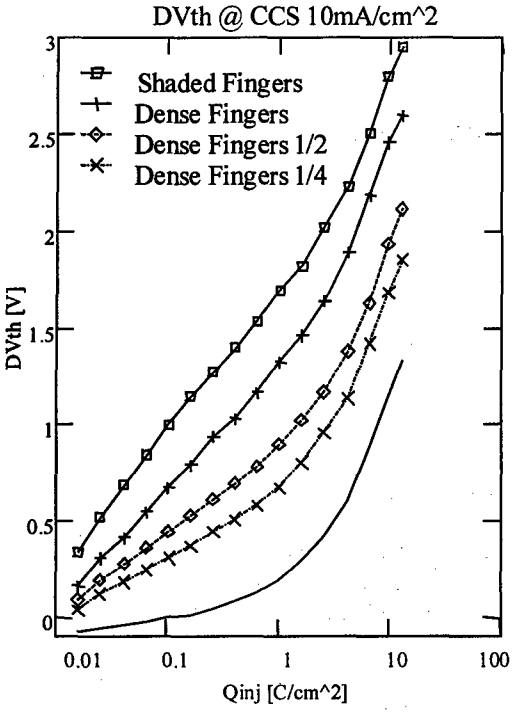



Fig. 3: Threshold voltage shift $\Delta V_{TH}$ as function of the injected charge $Q_{inj}$ (positive CCS stress).

Tab. 1: Geometrical parameters of the antennae considered. M.P.F. holds for Minimum Pitch Fingers. The Antenna Ratio (A.R.) is defined as the ratio between the antenna area and the gate area. The Minimum Size Antenna Ratio is due to the contact vias.

| Antenna Tipe | A. R. | Perimeter [μm] |
|---|---|---|
| Shaded M.P.F. | 1057.5 | 79702 |
| Dense Fingers M.P.F | 1057.5 | 79702 |
| Dense Fingers ½ M.P.F | 525.5 | 39590 |
| Dense Fingers 1/4 M.P.F | 263.3 | 7965 |
| Minimum Size | 1.5 | 80 |

A whole wafer has been tested to investigate the plasma damage impact on transistor characteristics and their dispersion over a 6" wafer. As an example, we show in fig.2 the threshold voltage dispersion before and after a CCS stress for a shaded-fingers M.P.F. antenna. The stress current density $J_{STRESS}$=33 mA/cm$^2$ was applied for a time $t_{STRESS}$=5s (total injected charge $Q_{Stress}$=165mC/cm$^2$). Before stress the threshold voltage is uniform across the wafer (<$V_{TH}$>=0.86V, $\sigma_{VTH}$=20mV), indicating that the plasma damage is only latent. After the CCS stress step the threshold voltage increased up to <$V_{TH}$>=1.94V. $V_{TH}$ strongly depends on the device position on wafer and the dispersion is noticeably

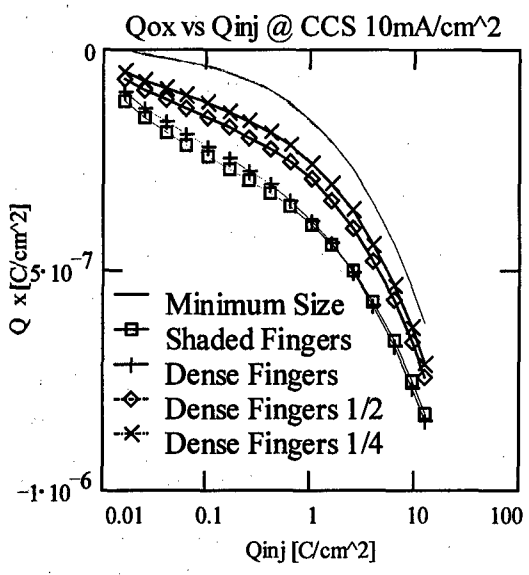



Fig. 4: Charge trapped $Q_{OX}$ during positive CCS stress as a function of the injected charge

grown ($\sigma_{VTH}$=740mV). Results reported in the following will compare devices with different antennae, but fabricated on the same site on the wafer surface. The maximum distance between adjacent antennae within each test structure is smaller than 1mm.

In order to depassivate the latent damage and minimize the newly stress induced damage, CCS tests have been performed with a stress current density $J_G$=±10mA/cm$^2$, that is, a stress condition less severe than those currently used in literature [3] (i.e., 400mA/cm$^2$). Between consecutive stress steps, we measured the MOSFET $I_D$-$V_G$ curve and the Fowler-Nordheim gate voltages $V_{FN}^+$ and $V_{FN}^-$ corresponding to a gate current density $J_G$=0.5mA/cm$^2$ and $J_G$=-0.5mA/cm$^2$, respectively. In this way, we have been able to monitor the threshold voltage (taken at the peak of the transconductance) and the oxide trapped charge, following the DiMaria's method [4]. The oxide charge is given by [4]:

$$Q_{ox}=C_{ox}(\Delta V_{FN}^+-\Delta V_{FN}^-) \quad (1)$$

where $C_{OX}$ is the gate oxide capacitance [F/cm$^2$] and $\Delta V_{FN}^{\pm}$ are the Fowler-Nordheim gate voltage shifts with respect to the values measured before CCS.

As shown in fig.3, during a positive stress $\Delta V_{TH}$ steadily increases with the injected charge $Q_{inj}$, indicating a dominant negative charge trapping.

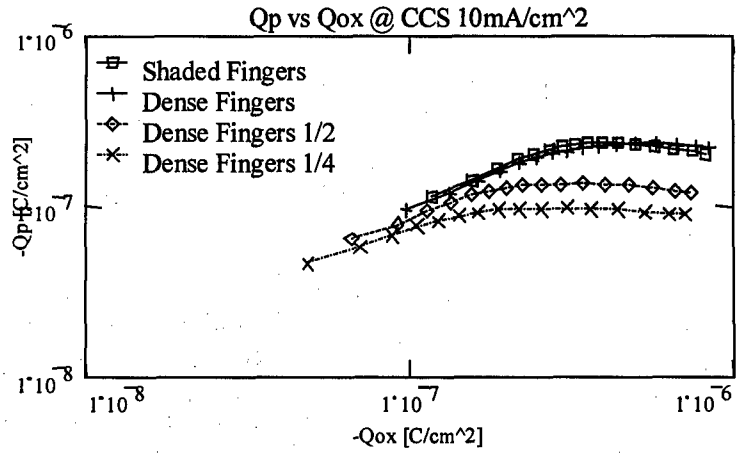



Fig. 5: $Q_P^{Antenna}$ as a function of $Q_{OX}$ for each antenna considered.

Only in the minimum-size devices, a positive charge is measured for low injected charge. Despite the large data dispersion, the oxide degradation mainly depends on the gate antenna, and mutual relationships between different antennae are preserved on each site of wafer surface.

In all the M.P.F. antennae the degradation is severe starting from the first stress step. Moreover, the higher the antenna ratio, the higher the oxide trapped charge. After stress, the Minimum Size MOSFET presents the minimum degradation and dispersion, as shown in tab. 2. For negative stresses, the same results hold.

Tab. 2: Minimum Size $\Delta V_{TH}$ dispersion as a function of injected charge during the positive stress.

| Injected Charge | $\sigma_{\Delta VTH}$ |
|---|---|
| as received | 16mV |
| 0.1C/cm$^2$ | 50mV |

Tab. 3: Minimum Size $Q_{OX}$ dispersion as function of injected charge during the positive stress.

| Injected Charge $Q_{inj}$ | $\sigma_{QOX}$ |
|---|---|
| 0.16 [C/cm$^2$] | -4.17nC/cm$^2$ |
| 1.6 [C/cm$^2$] | -9.28nC/cm$^2$ |

We have investigated the different trapping behavior of the oxide for each antenna [4] and for both positive and negative stress polarity. In fig.4 we reported the trapped charge $Q_{OX}$ for all the antennae as a function of the injected charge $Q_{inj}$ during the positive stress. The same results hold for the negative stresses. As in case of $\Delta V_{TH}$, the minimum size presents minimum $Q_{OX}$ dispersion and degradation across the wafer during both positive and negative stresses (tab. 3).

## 3. A new approach to evaluate the plasma damage

The main "signatures" of plasma damage are the antenna dependent oxide degradation [1,5] and the position dependent degradation and dispersion [5]. Since the Minimum Size presents minimum degradation and negligible dispersion across the wafer (see tab. 2 and 3), we take it as the "reference" device, considered as plasma-damage-free. In other words, in Minimum Size devices we suppose no latent damage present and no depassivation effect activated by CCS. Hence, by comparing each antenna with the minimum size, we are able to give a measurement of stress induced depassivation of plasma damage. Since the minimum-size is plasma damage free, for each stress step the difference between the charge trapped in each antenna and the minimum-size gives immediately a meter to evaluate the plasma damage depassivation effect. In detail: we have introduced the plasma-related charge $Q_P^{Antenna}$ defined as

$$Q_P^{Antenna} = Q_{OX}^{Antenna} - Q_{OX}^{Minimum\ Size} \quad (2)$$

where $Q_{OX}^{Antenna}$ is the oxide charge (eq. 1) for a given antenna [4]. At each stress step, $Q_P^{Antenna}$ accounts for latent damage depassivation.

In order to investigate the dependence of plasma damage from geometry and area, in fig. 5

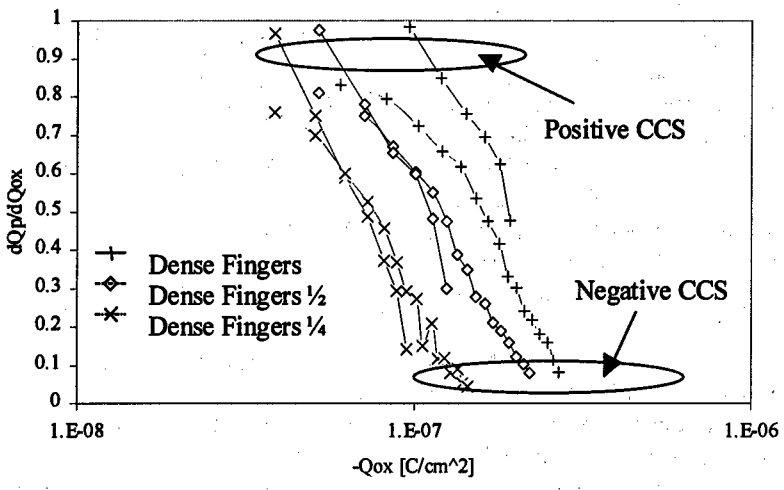



Fig. 6: Stress depassivation kinetic $\partial Q_P^{Antenna}/\partial Q_{OX}$ vs $Q_{OX}$ for positive (solid) and negative (dashed) CCS early stress steps.

(positive CCS) we reported $Q_P$ versus $Q_{OX}$ for all the antennae. At low $Q_{OX}$, $Q_P$ steadily increases, while at high $Q_{OX}$ $Q_P$ saturates at a value characteristic of the antenna (and linked to the position on the wafer). In fact, at low $Q_{OX}$ the contribution of plasma damage is the dominant effect. At high $Q_{OX}$, starting from $Q_{OX} \approx -5*10^{-7}$ C/cm$^2$ the contribution of the plasma-induced charge depassivated during CCS becomes negligible and any further charge trapped is due only to newly CCS induced defects. The same results hold for the negative CCS. The depassivation kinetic of plasma damage during negative CCS stress is slower than that during the positive CCS: the latent damage depassivation kinetics for low trapped charge is given plotting $\partial Q_P^{Antenna}/\partial Q_{OX}$ versus $Q_{OX}$. As shown in fig. 6, the initial latent damage depassivation is a strong function of the antenna connected to the gate. The higher the antenna ratio, the higher the initial charge trapped during the first stress step. The depassivation kinetics is a function of the stress polarity: during the positive CCS $\partial Q_P^{Antenna}/\partial Q_{OX} \approx 0.9$, while during the negative CCS $\partial Q_P^{Antenna}/\partial Q_{OX} \approx 0.75$, indicating for the negative CCS stress a plasma damage depassivation kinetics slower than during the positive one.

Regardless the stress polarity, the $Q_{PSAT}$ saturation value is a meter of the induced damage at the end of the manufacturing process. The $Q_{PSAT}$ value strongly depends on the device antenna type and position on the wafer. For a chosen antenna area, the plasma damage depends on the antenna geometry: as shown in tab. 4, $Q_{PSAT}$(Shaded) is slightly higher than $Q_{PSAT}$(Dense-Fingers) due to electron shading effect [6]. For a chosen dense-fingers antenna geometry, $Q_{PSAT}$ gives a damage dependence from the area: $Q_{PSAT}$(Dense-Fingers) is roughly twice as $Q_{PSAT}$ (Dense-Fingers ½) and three times $Q_{PSAT}$ (Dense-Fingers 1/4) [1].

Tab. 4: $Q_{PSAT}$ values for the antennae considered (positive CCS)

| Antenna | $Q_{PSAT}$ [C/cm$^2$] |
|---|---|
| Shaded Fingers | $-2.45*10^{-7}$ |
| Dense Fingers | $-2.4*10^{-7}$ |
| Dense Fingers 1/2 | $-1.36*10^{-7}$ |
| Dense Fingers 1/4 | $-9.75*10^{-8}$ |

These results propose a very useful characterization technique in an automated probing environment for accelerated reliability characterization of different test structures. After the characterization of the stress level ($Q_{OX}$) necessary to saturate $Q_P$ for a given manufacturing technology, the $Q_{PSAT}$ value can be measured with a single CCS stress step to the given $Q_{OX}$ (equivalently $Q_{inj}$) in order to completely remove the effects of depassivation. Mapping $Q_{PSAT}$ (or $\Delta V_{TH}$) on the wafer surface permits to investigate the reliability impact of plasma damage for different technologies, machines, processes, with an easy extension to statistical analysis of the lot-to-lot dispersion.

## 4. Conclusions

During the stress, the antenna effect plays a key role in determining the plasma damage. In fully

processed structures, we have been able to determine the oxide damage due to depassivation of plasma-induced defects by performing constant current stress and measurements of the main MOS parameters. Plasma damage has been quantitatively evaluated by comparing the electrical characteristics of MOS featuring various Antenna Ratios with a reference minimum-size device. The device has been considered as plasma-damage-free in our work, as it shows minimum modifications and small dispersion of the electrical parameters after stress.

We have introduced an original description of the kinetics of plasma damage as revealed by CCS, featuring a two-slope curve. At low injected charge, the depassivation of plasma damage is the main responsible of the oxide charge increase. The growth kinetics of the plasma-related oxide charge depends from the antenna characteristics. At high injected charge, the oxide charge relative to plasma damage reaches a saturation value, which depends on the antenna characteristics as well.

Hence the measured oxide trapped charge [4] can give a quantitative measurement of the net plasma induced defects depassivated during stress. This experimental method can be easily extended to accelerated reliability characterization of test structures and immediately gives the optimal stress condition for plasma damage evaluation.

## Acknowledgements

This work has been partially supported by SGS-Thomson Microelectronics under contract MEDEA T502 "Options for 0.35μm CMOS". The first author would like to thank M.Barbazza for his invaluable help.

## References

[1] H.Shin, C.Hu. IEEE T. on Sem. Man. 6(1993), 96.

[2] O.O. Awadelkarim, S.J. Fonash, P.I. Mikulan, Y.D. Chan. J. Appl. Phys. 79(1996), 517.

[3] K.P. Cheung. 1st Symp. Plasma Process-Induced Damage, 1996, 11-14.

[4] D.J. DiMaria. J. Appl. Phys. 47(1976), 4073.

[5] J.P.McVittie. 1st Symp. Plasma Process-Induced Damage, 1996, 7-10.

[6] K.Hashimoto, Y.Hikosaka, A.Hasegawa. M.Nakamura. 1st Symp. on plasma Process-Induced Damage, 1996, 43-46.

[7] T.Brozek, L.Peng and C.R.Viswanathan. 1st Symp. Plasma Process-Induced Damage, 1996, 81-83.

PERGAMON

Microelectronics Reliability 38 (1998) 925–929

MICROELECTRONICS RELIABILITY

# Low frequency noise analysis as a diagnostic tool to assess the quality of 0.25$\mu$m Ti-silicided poly lines

E.P. Vandamme [a], I. De Wolf [a], A. Lauwers [a] and L.K.J. Vandamme [b]

[a] *IMEC, Kapeldreef 75, B-3001 Leuven, Belgium, fax. −32−16−281 214*
[b] *Eindhoven University of Technology, Dept. of El. Eng., EH 5.15, 5600MB Eindhoven, The Netherlands*

**Abstract**

In this paper, dc and low frequency noise measurement results on Ti-silicided poly lines are presented and analysed. Besides Raman scattering [1], low frequency noise analysis gives additional information on the presence and the spatial distribution of the high resistive C49 phase in the silicide line. The low frequency noise is strongly dependent on whether or not the distribution of the C49 phase is uniform or spotted-like. The experimental results agree with our model for a uniform C49 phase distribution over the silicided line. A spotted-like distribution of the C49 phase, which causes non-uniform current densities, can not explain the observed noise behaviour. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Titanium silicides are widely used in CMOS, BiCMOS, NVM and other silicon-based technologies to lower, a.o. the short-range interconnect resistance of poly-silicon lines. Although $CoSi_2$ offers better scalability towards very deep submicron technologies [2], $TiSi_2$ is still widely used down to 0.25$\mu$m technologies. During the silicidation process, a high resistive crystalline phase is formed, called C49. The resistivity of this phase is approximately 60 to 100$\mu\Omega$cm. During thermal annealing, the C49 phase is transformed into a C54 phase, with lower resistivity of about 14$\mu\Omega$cm [3]. However, for narrow, thin lines, incomplete transformation may occur during thermal annealing steps and the C49 phase can still be present in the $TiSi_2$-line [4]. This increases the total resistance of the poly-line, which is detrimental for good contact formation. For analog applications, extra noise will be induced in the circuit, e.g. at the input of an amplifier [5], due to the presence of the C49 phase. This deteriorates circuit performance. It was shown that $\mu$Raman scattering [1] provides information on the spatial distribution of the C49 phase. A drawback of $\mu$Raman scattering is that the spot size is larger than the width of the poly line. As a result, only information with a spatial resolution of 1$\mu$m can be extracted. In this paper we show that additional information on the spatial distribution of the C49 phase can be extracted from $1/f$ noise measurements. Low frequency noise experiments can distinguish between a uniform distribution of the C49 phase across the

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00114-0

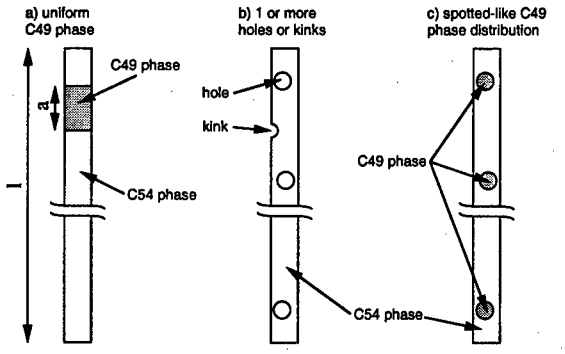



Fig. 1. *a)* Uniform C49 phase distribution. *b)* Uniform C54 phase with kinks and holes. *c)* Spotted-like C49 phase distribution.

silicided line or a spotted-like distribution that causes current crowding.

## 2. Noise modelling

In this section $1/f$ noise models are discussed for *a)* a uniform C49 phase distribution in a silicided poly line, *b)* a pure C54 phase suffering from holes and kinks and *c)* a spotted-like distribution of the high resistive C49 phase, as illustrated in Figures 1a, b and c, respectively. The uniform C54 phase distribution with kinks and holes can be seen as a limiting case for the spotted-like C49 phase distribution. Both distributions result in non-uniform current densities.

### *2.1. Uniform C49 phase distribution*

The resistance for this type of structure is given by the following equations

$$R_{54} = \rho_{54}\frac{l-a}{w} \tag{1}$$

$$R_{49} = \rho_{49}\frac{a}{w} \tag{2}$$

$$R = R_{54} + R_{49} = R_{\min} + (\rho_{49} - \rho_{54})\frac{a}{w} \tag{3}$$

with $w$ and $l$ the width and total length of the silicided poly line, $a$ the total length of the C49 phase and $\rho_{49}$, $\rho_{54}$ the sheet resistance of the C49 and C54 phase, respectively. The value $R_{\min} = \rho_{54}l/w$ denotes the resistance of the silicided poly line if no C49 phase is present, i.e. pure C54. The resistance noise power spectral density can be calculated as [6]

$$S_{R_{54}} = \frac{C_{54}R_{54}^2}{f(l-a)w} = \frac{C_{54}\rho_{54}^2}{fw^3}(l-a)$$

$$= K_{54}(l-a) \tag{4}$$

$$S_{R_{49}} = \frac{C_{49}R_{49}^2}{faw} = \frac{C_{49}\rho_{49}^2}{fw^3}a = K_{49}a \tag{5}$$

$$S_R = S_{R_{54}} + S_{R_{49}} = S_{R_{\min}} + (K_{49} - K_{54})a \tag{6}$$

with $C_{49}$ and $C_{54}$ the $1/f$ noise parameter in the C49 and C54 phase, respectively and $f$ the frequency. The value $S_{R_{\min}} = C_{54}\rho_{54}^2 l/fw^3$ denotes the power spectral density of the silicided line if no C49 phase is present, i.e. pure C54. Eqs. 3 and 6 show that both $R$ and $S_R$ are proportional to $a$. Note that the sole variable in this system is $a$, the total length of the C49 phase. Rearranging Eqs. 3 and 6 yields

$$S_R - S_{R_{\min}} \propto R - R_{\min} \tag{7}$$

Thus, for a uniform C49 phase distribution, a linear relationship exists between extra noise caused by the C49 phase, $S_R - S_{R_{\min}}$, and additional resistance, $R - R_{\min}$.

### *2.2. Uniform C54 phase with kinks and holes*

As a limiting case for a spotted-like C49 phase distribution, we describe the effect of kinks and holes in a silicided poly line on the resistance and resistance noise power spectral density. Based on the analysis in [7] for 1 hole or kink, the equations for resistance and noise can be extended to multiple holes. To account for the influence of holes and kinks, the silicided poly line is divided into a part with uniform C54 phase, having total length $l - xw$ and width $w$, and a part with holes or kinks, having total length $xw$ and width $w_{\text{eff}}$. The total number of holes or kinks is represented by the variable $x$. The effective width $w_{\text{eff}}$ of the silicided poly line is smaller in presence of a hole or kink, thus $w_{\text{eff}} < w$. The resistance of that part of the silicided poly line with kinks and holes equals

$xR_h$. For this type of structure, the resistance can be calculated using the following equations

$$R_{54} = \rho_{54}\frac{l - xw}{w} \tag{8}$$

$$R_h = \rho_{54}\frac{w}{w_{\text{eff}}} \tag{9}$$

$$R = R_{54} + xR_h = R_{\min} + \rho_{54}x\left[\frac{w}{w_{\text{eff}}} - 1\right] \tag{10}$$

Note that the effect of current crowding is taken into account by the variable $w_{\text{eff}}$. The resistance noise power spectral density for this type of structure is given by the following equations

$$S_{R_{54}} = \frac{R_{54}^2 C_{54}}{fw(l - xw)} \tag{11}$$

$$S_{R_h} = \frac{R_h^2 C_{54}}{fww_{\text{eff}}} \tag{12}$$

$$S_R = S_{R_{54}} + xS_{R_h}$$

$$= S_{R_{\min}} + x\frac{\rho_{54}^2 C_{54}}{fw^2}\left[\left(\frac{w}{w_{\text{eff}}}\right)^3 - 1\right] \tag{13}$$

Since the noise sources are uncorrelated, the total resistance noise power spectral density is the summation of the contributions from the C54 part and the part with kinks or holes. Thus, for the additional resistance and noise power spectral density, we can derive the following equations

$$R - R_{\min} \propto x\left[\left(\frac{w}{w_{\text{eff}}}\right) - 1\right] \tag{14}$$

$$S_R - S_{R_{\min}} \propto x\left[\left(\frac{w}{w_{\text{eff}}}\right)^3 - 1\right] \tag{15}$$

Note that in this model, both the number of holes or kinks as well as the effective width of the poly line are variables.

### *2.3. Spotted-like C49 phase distribution*

As was stated earlier, the distribution with kinks and holes can be seen as a limiting case for a spotted-like C49 distribution. However, the resistivity of the C49 phase is not infinite like for a hole and some modifications must be made to the previous model. This can be easily done by assuming a different effective width of the poly-line, $w_{\text{eff}}^*$ instead of $w_{\text{eff}}$, as used in Eqs. 8–15, with $w_{\text{eff}}^* > w_{\text{eff}}$. The variable $w_{\text{eff}}^*$ takes into account the resistivity of the C49 phase. The same conclusions can then be drawn for a spotted-like C49 distribution as for a uniform C54 phase with kinks or holes.

## 3. Measurement set-up

Figure 2a shows a cross section of the sample. A poly-silicon layer of 250nm was deposited on 150nm oxide. After poly patterning, oxide spacers were formed. The $TiSi_2$ was formed by sputtering 55nm of Ti followed by a Rapid-Thermal Anneal (RTA) at 730°C, a selective etch and a second RTA at 850°C. The final silicide thickness is about 70nm. The experimental set-up is shown in Figure 2b. A four-terminal structure is used. The length of the poly line between $C$ and $D$ is 150$\mu$m. The width is 0.25$\mu$m.

The noise measurements were performed on wafer level. A constant current is passed through the DUT using a battery with a series resistance $R_V$, which is at least 20 times larger than the resistance of the silicided poly line. The voltage noise power spectral density $S_V$ is measured at constant current $I$ using a low-noise voltage preamplifier, BD5004, and an FFT spectrum analyser, HP35665A. The resistance noise power spectral density is calculated using $S_V/V^2 = S_R/R^2$ [6]. To correct for thermal noise stemming from the DUT, the bonding pads and the connecting wires, the background noise spectrum is measured with $I = 0$, i.e. $S_V$ with $V_b$ replaced by a short circuit.

## 4. Results and discussion

A typical background and $1/f$ noise spectrum is shown in Fig. 3. As stated in the section on the different noise models, the difference between measured and minimum resistance noise power spectral density should be plotted as a function of its difference in resistance. The minimum resistance of the poly line when no C49 phase is present, $R_{\min}$ was obtained from 4-point resistance measurements on more than 100 devices and

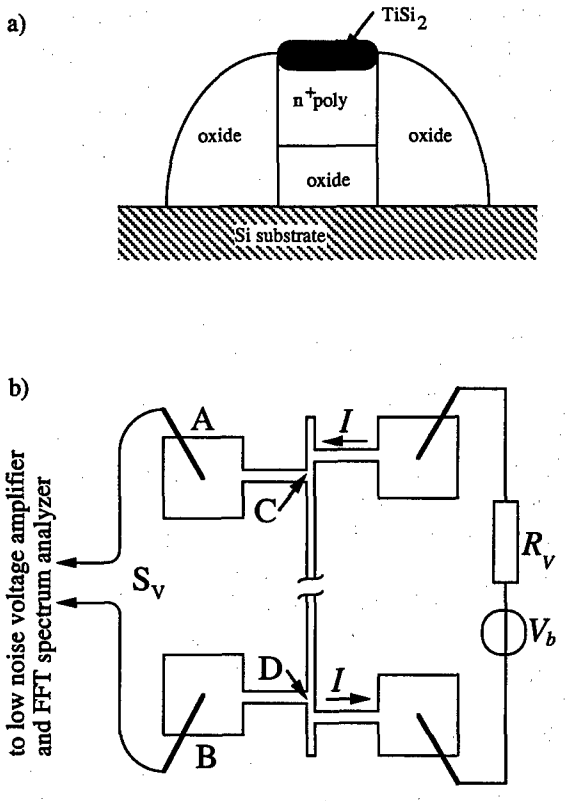



Fig. 2. a) Device cross-section. b) Noise measurement set-up

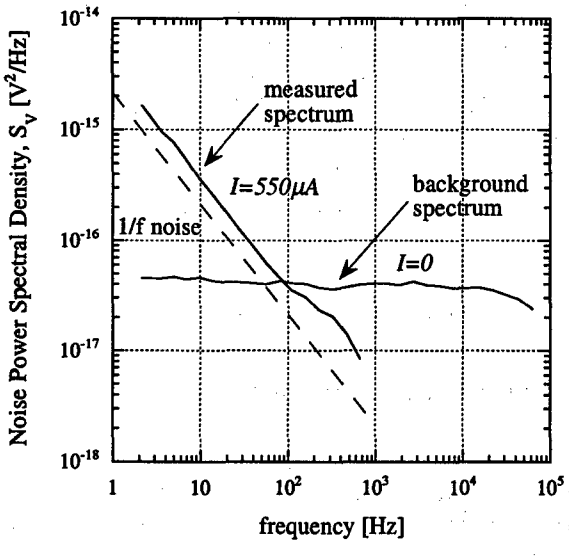



Fig. 3. Typical noise spectrum measured at $I = 0$ (background) and $I = 550\mu$A, with background correction. The background noise level corresponds to a resistance of 2.5kΩ, which is in agreement with a 2-point dc resistance measurement between A-B (see Fig. 2b).

equals 900Ω. The calculated sheet resistance of the C54 phase is 1.5Ω/□, a typical value for this type of silicide. The corresponding measured resistance noise power spectral density, $S_{R_{\min}}$ equals $1.3 \times 10^{-10}\Omega^2/\text{Hz}$ at $f = 1$Hz.

For a spotted-like C49 phase distribution across the silicided poly line, two distinct situations should be analysed: *i)* fixed number of C49 spots with variable effective poly width and *ii)* variable number of C49 spots with fixed effective poly width. In case the effective width of the poly line is constant one expects a linear relation between additional noise and resistance, since both $S_R - S_{R_{\min}}$ and $R - R_{\min}$ are proportional to $x$, the number of C49 spots. For a fixed number of C49 spots, one should distinguish following situations: *i)* $w_{\text{eff}}$ much smaller than $w$, i.e. relatively large spots and *ii)* $w_{\text{eff}} \approx w$, i.e. very small spots and minor current crowding. In the former case we expect the noise to vary with the third power of the difference in resistance, i.e. $S_R - S_{R_{\min}} \propto (R - R_{\min})^3$, if $x$ is constant. If $w_{\text{eff}} \approx w$, a linear relation between $S_R - S_{R_{\min}}$ and $(R - R_{\min})$ is expected using a Taylor expansion with $\lim_{\Delta w \to 0}(w_{\text{eff}} = w - \Delta w)$.

If the C49 phase is uniformly distributed across the poly line, we expect a linear dependence of the additional resistance noise versus its difference in resistance, i.e. $S_R - S_{R_{\min}} \propto R - R_{\min}$ according to Eq. 7.

Figure 4 shows the experimental dependence of the additional resistance noise power spectral density versus its difference in resistance for different samples. A linear relationship between $S_R - S_{R_{\min}}$ and $R - R_{\min}$ is observed. Note that $R$ and $S_R$ stem from section C to D of the poly line, as shown in Fig. 2b.

For a spotted-like C49 phase distribution, a linear dependence of $S_R - S_{R_{\min}}$ versus $R - R_{\min}$ can only be expected if on the one hand $w/w_{\text{eff}}$ is constant and $x$ varies or on the other $w_{\text{eff}} \approx w$ and $x$ is constant. However, when calculating absolute numbers for $R$, $R_{\min}$, $S_R$ and $S_{R_{\min}}$ using realistic values for $\rho_{49}$, $\rho_{54}$, $C_{49}$ and $C_{54}$, the spotted-like C49 distribution cannot explain the observed linear dependence. For a fixed $x$ and varying small $w_{\text{eff}}$, the predicted change in total resistance and noise of the poly line due to the presence of C49 spots is much too low compared to measurement

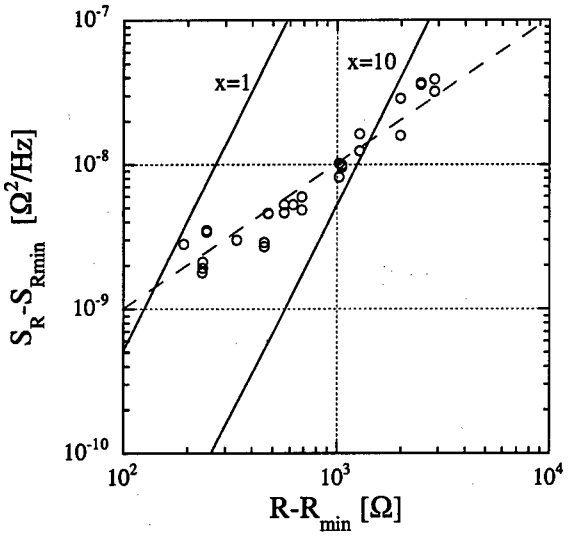



Fig. 4. Difference in measured and minimum resistance noise power spectral density at 1Hz as a function of its difference in resistance. The circles denote measurement results, the dashed line has a slope 1, see Eq. 7. The full lines are calculations based on the non-uniform current density approach with varying $w_{\text{eff}}$ and 1 or 10 holes, see Eqs. 8–15.

results, i.e. $S_R - S_{R_{\min}} \ll 1 \times 10^{-10}\Omega^2/\text{Hz}$ at $f = 1\text{Hz}$ and $R - R_{\min} \ll 100\Omega$. On the other hand, for a fixed $w_{\text{eff}}$ and varying $x$, calculated values for $S_R - S_{R_{\min}}$ are close to measurement results. However, the predicted change in total resistance due to the presence of C49 spots is more than 2.5 times too low compared to measurement results. Besides, the large number of spots required to calculate the difference in resistance is not in agreement with Figure 5 in [1] and conclusions therein where a local occurrence of the C49 phase was observed.

Based on the preceding analysis, it is now clear that, from the models presented before, we can conclude that the distribution of the C49 phase is uniform across the silicided line. A spotted-like distribution of the C49 phase, as shown in Fig. 1c), can not explain our experimental results.

## 5. Conclusions

Low frequency noise measurements were performed on $TiSi_2$ poly lines. We showed that noise spectroscopy provides additional information on the microscopic structure of silicided poly lines. Pure $1/f$ noise was observed. From the observed linear dependence of the noise power spectral density $S_R$ on resistance, we can conclude that the distribution of the occurring C49 phase is uniform across the silicided line. A spotted-like distribution of the C49 phase, causing non-uniform current densities, can not explain the experimental results.

## Acknowledgements

The work of E.P. Vandamme is supported by IWT. The authors would like to thank D. Howard and K. Maex for providing the samples.

## References

[1] De Wolf I et al. A reliability study of Titanium Silicide lines using micro-Raman spectroscopy and emission microscopy. Microelectron. Reliab. MR-37 (1997) 1591–1594.

[2] Maex K. $CoSi_2$: an attractive alternative to $TiSi_2$. Semicon. Int. march (1995) 75–80.

[3] Nicolet M and Lau S. VLSI Electronics Microstructure Science 6 (1983) p 415.

[4] Ganin E et al. $TiSi_2$ formation on submicron polysilicon lines: role of line width and dopant concentration. Rapid thermal and int. proc. II. MRS proc. (1993) 109–114.

[5] Laker K and Sansen W. Design of analog integrated circuits and systems. McGraw-Hill, 1994.

[6] Vandamme L. Noise as a diagnostic tool for quality and reliability of electronic devices. IEEE Trans. Electron Devices. TED-41 (1994) 2176–2187.

[7] Vandamme L and van Kemenade A. Resistance noise measurement: a better diagnostic tool to detect stress and current induced degradation. Microelectron. Reliab. MR-37 (1997) 87–93.

PERGAMON

Microelectronics Reliability 38 (1998) 931–936

MICROELECTRONICS RELIABILITY

# Hot carrier degradation mechanisms in sub-micron *p* channel MOSFETs: Impact on low frequency (1/f) noise behaviour

E. Sheehan, P.K. Hurley and A. Mathewson

National Microelectronics Research Centre, University College Cork,
Lee Maltings, Prospect Row, Cork, Ireland.

**Abstract**

In this work we present new results which illustrate the impact of hot carrier (HC) degradation on the low frequency (1/f) noise behaviour of submicron *p* channel MOSFETs. Submicron *p* channel MOSFETs were subjected to HC stress at a range of gate bias conditions, and the response of the low frequency noise was recorded. The results obtained are in marked contrast to the reported influence of HC stress on *n*MOSFETs 1/f noise, and indicate that the measurement of 1/f noise is a useful tool for investigating HC induced aging effects in submicron *p* channel devices. The significance of these results to the use of *p*MOSFETs in analog applications is briefly discussed.
© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Several previous studies have reported that hot carrier (HC) degradation [1-4] and plasma charging damage [5] have a significant impact on the low frequency (1/f) noise behaviour of *n* channel MOSFETs. The increase in the 1/f noise levels has been attributed to a generation of slow states in the gate oxide during the high field stress. Analysis has identified the generated slow states to be located in a narrow region (<50 nm) surrounding the drain of the device [1,2]. Furthermore, the sensitivity of 1/f noise to the Si-$SiO_2$ interface properties has resulted in the use of 1/f noise measurements as a monitor for HC induced oxide degradation in *n* channel devices.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00143-7

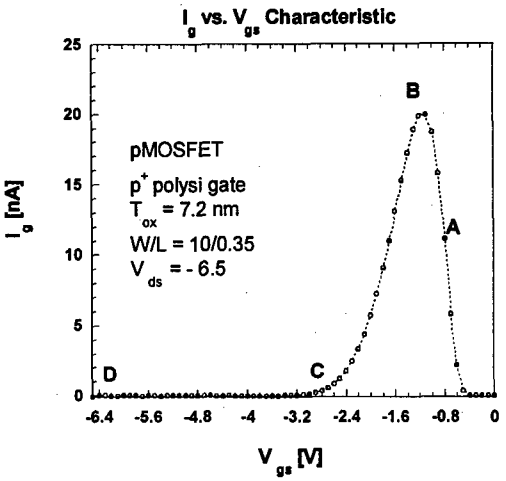



Figure 1: Typical gate current ($I_g$) versus gate voltage ($V_{gs}$) characteristic of the *p*MOSFETs used in this work. The points A, B, C and D represent the different stress conditions that were examined in order to assess their impact on the *p*MOSFET 1/f noise.

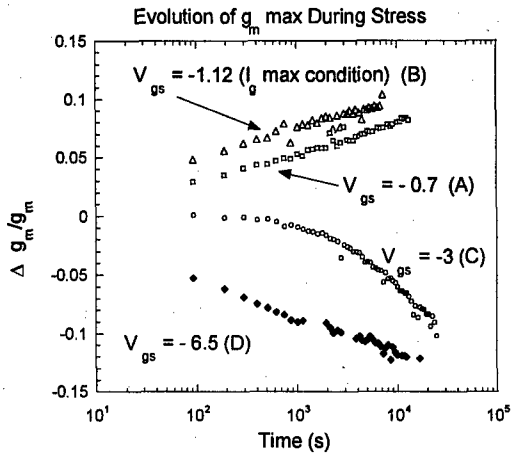



Figure 2: The evolution of the maximum transconductance during hot carrier stress at the bias conditions A, B, C and D identified in Figure 1. *p*MOSFET, W/L = 10/0.35.

However, in a recent work [6] it was demonstrated that in the case of *p* channel MOSFETs the 1/f noise was unaffected by HC degradation, for devices stressed at the condition of the maximum gate current. This previous work was performed on buried channel *p*MOSFETs with minimum channel lengths in excess of one micron. The purpose of this work is to extend the previous studies to investigate surface channel *p*MOSFETs in the submicron regime. Furthermore, the work investigates the response of *p*MOSFET 1/f noise to HC stressing over a range of gate bias ($V_{gs}$) conditions, from $V_{gs} \approx V_t$ to $V_{gs}=V_{ds}$. The results are interpreted in relation to the various mechanisms of degradation occurring in submicron *p* channel devices [7]. The significance of these new results is briefly discussed.

## 2. Experimental Results

The experimental results were obtained for a range of submicron CMOS processes incorporating both *n*+ and *p*+ polysilicon gates for the *p* channel devices. The low frequency noise was measured using a custom made low noise amplifier and a HP35665A dynamic signal analyzer. All the 1/f noise measurements were performed in the linear region of device operation, as it is known that for devices biased in the saturation region, the degraded region of the device is shielded by the pinch-off region around the drain [1].

Figure 1 presents a typical gate current versus gate bias characteristic for the devices measured (see Figure for details). The points A, B, C and D represent the different stress conditions that were examined in order to assess their impact on the *p*MOSFET 1/f noise. The HC stress measurements were periodically interrupted to determine variation in the linear region threshold voltage and the maximum transconductance ($V_{ds}$ = -100 mV). Figure 2 shows the evolution of the maximum transconductance ($g_m$ max) corresponding to stress conditions A, B, C and D. The pattern of the $g_m$max degradation results are in broad agreement with those obtained by Woltjer et al. [7].

Figure 3 illustrates that for devices stressed at the maximum gate current ($I_g$ max) condition (region B in Figures 1 and 2), the 1/f noise exhibits no

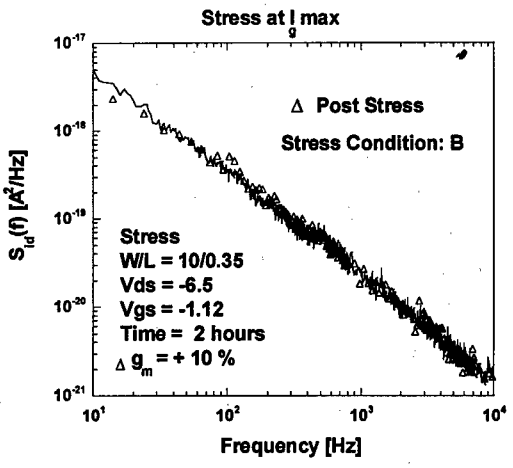



Figure 3: Drain current noise spectral density before and after hot carrier stress for *p*MOSFET (W/L=10/0.35) stressed at $I_g$ max. The stress conditions are shown in the inset.

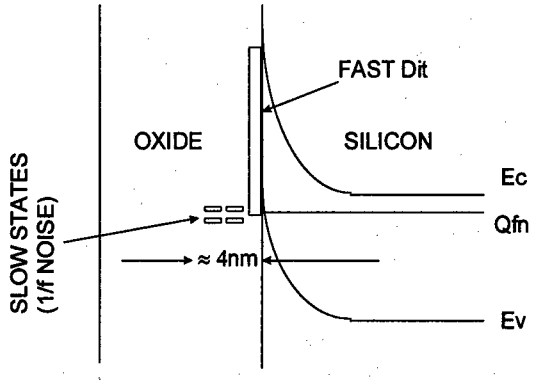



Figure 4: Schematic energy band diagram for the *p*MOSFET in inversion. The figure illustrates that only slow states at energies corresponding to Ev contribute to low frequency (1/f) noise.

measurable change, even after significant levels of device degradation.

This result has been confirmed on submicron *p*MOSFETs from alternative CMOS processes incorporating n+ polysilicon gates for the *p* channel devices. The result has also been obtained for a range of channel lengths. In addition, hot carrier degradation at stress conditions between B and A (see Figures 1 and 2), also have no impact on the 1/f noise of the device. This confirms the results in [6], and extends the previous work to demonstrate that for surface channel sub-micron *p*MOSFETs (0.35μm), the magnitude of the 1/f noise is not influenced by HC stressing at the condition of the maximum gate current. It is noted that these results are in marked contrast to the case of *n* channel MOSFETs, where large increases in 1/f noise occur for devices subjected to HC stressing [1-4,8].

The insensitivity of the *p*MOSFET 1/f noise to the HC stress (at $I_g$ max) can be understood with reference to Figures 4 and 5. MOSFET 1/f noise is generally agreed to originate from an interaction of the channel inversion charge with slow states located in the gate oxide [9]. Hence, for *p* channel devices, only slow states at energies corresponding to the valence band edge ($E_v$) in the silicon can contribute to the measured 1/f noise. This is illustrated in Figure 4. For an increase in 1/f noise to be measured after HC stress, it is required that additional slow states are generated in the gate oxide (near $E_v$), *and that the generated slow states can interact with the channel inversion charge*, via a tunneling process. For stress at $I_g$ max, electron trapping in the gate oxide results in an effective extension of the drain region [10]. This extended drain region masks any possible interaction of generated slow states with the channel inversion charge (see Figure 5). Hence, no increase in 1/f noise is measured for this stress condition.

As a consequence of the drain region extension, for stress at the Ig max condition, it is not possible to assess if slow state generation is occurring in the gate oxide. However, if *p* channel devices were stressed at a bias condition which exposes the degraded region of the gate oxide, then measurable changes in the device 1/f noise should be observed after the HC stress, *if slow states are being generated during the oxide degradation.* From Figure 2, for devices stressed at bias condition C, $g_m$ max is observed to decrease during the HC stress.

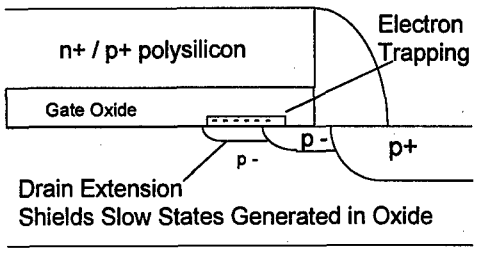



Figure 5: Schematic cross-section through a *p*MOSFET structure. The Figure illustrates that electron trapping in the gate oxide, for stress at Ig max, extends the drain region shielding the channel from generated interface/slow states.

Under these stress bias conditions the device degradation is due to interface state generation by holes, and the degraded region is not shielded by an extension of the drain [7]. Hence, if slow states are generated in the gate oxide, 1/f noise variation should be measurable for *p*MOSFETs stressed in this bias regime. This is confirmed in Figure 6, which shows a significant increase in the device 1/f noise after stress. This result was reproducible across the wafer, and for *p*MOSFETs from an alternative CMOS process. Figure 7 shows the evolution of the 1/f noise (at 80Hz) in the linear region during the stress at bias point C. The evolution of the 1/f noise at other frequencies in the 1/f region (e.g., 20 Hz, 320Hz), exhibited the same trend. The plot indicates an approximately linear increase in the 1/f noise with stress time.

The results in Figures 6 and 7 demonstrate that for stress at bias condition C, slow states are generated in the gate oxide in addition to the interface states detected by the charge pumping technique [7,11]. An exact determination of the increase in the slow state density is complicated, as the degraded area of the gate oxide is not precisely known. However, it is clear that at bias point C, the device 1/f noise is a far more sensitive monitor of the device degradation than $g_m$ max. An increase in the 1/f noise of ≈400% occurred for a decrease of ≈13 % in $g_m$ max.

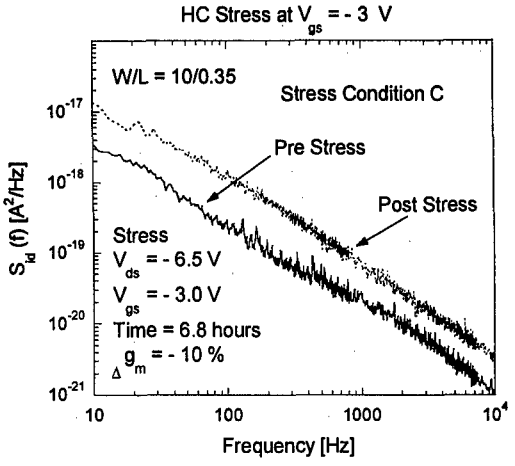



Figure 6: Drain current noise spectral density before and after hot carrier stress for pMOSFET (W/L=10/0.35) stressed at $V_{GS}$ = -3 (C in Figure 1). The stress conditions are shown in the inset.

Further results have been obtained which investigate *p*MOSFET device degradation at the condition $V_{gs} = V_{ds}$. For *p*MOSFET stress at this bias condition, the precise mechanism of device degradation has been in dispute. Some works attribute the decease in $g_m$ max to the generation of interface states [12], where other works identify positive charge generation (hole trapping) as the degradation mechanism [13]. The measurement of 1/f noise at the stress condition $V_{gs} = V_{ds}$ provides a means of further investigating the degradation mechanism. If the degradation results solely from hole trapping in the gate oxide, then no increase in the device 1/f noise is expected. However, if the degradation is a result of interface state generation, then, (as at bias point C), it is expected that the 1/f noise in the linear region of device operation will increase with the device degradation.

The results in Figure 8 show the measured 1/f noise before and after HC stressing at the condition D ($V_{gs}$ = $V_{ds}$). The HC stress resulted in a transconductance degradation of –10 %. However, it is evident from Figure 8 that the device 1/f noise does not vary with the stress in this bias region. This result was reproducible across the wafer. The results in Figure 8 support hole trapping in the gate oxide as the mechanism of device degradation at the stress condition $V_{gs} = V_{ds}$ for *p*MOSFETs.

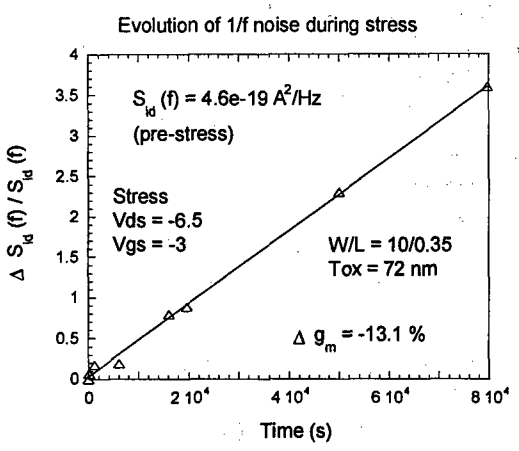



Figure 7: Evolution of 1/f noise at 80 Hz during stress time for *p* channel MOSFET stressed at bias point C. The data is expressed as the percentage increase in the 1/f noise.

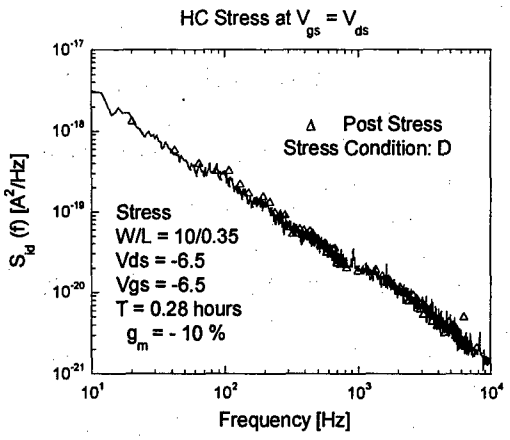



Figure 8: Drain current noise spectral density before and after hot carrier stress for pMOSFET (W/L=10/0.35) stressed at $V_{gs} = V_{ds}$ (D in Figure 1). The stress conditions are shown in the inset.

## 3. Conclusion and Significance

In conclusion, new results have been presented demonstrating the impact of HC degradation on the 1/f noise of submicron surface channel *p*MOSFETs. For device stress at the condition of the maximum gate current, the 1/f noise is not affected by the device degradation due to the degraded region of the gate oxide being shielded by an extension of the drain region. For stress at more negative gate voltages ($V_{gs} \approx V_{ds}/2$), where the damaged region of the gate oxide is not shielded from the channel region, the device 1/f noise does increase during the stress. These results demonstrate that slow states are generated in the gate oxide in addition to interface states at this stress condition for *p* channel devices. For *p*MOSFET stress at $V_{gs} = V_{ds}$, the 1/f noise is again insensitive to the HC stressing. This result supports hole trapping in the oxide as the degradation mechanism at this stress condition.

A number of significant issues arise from these observations. Firstly, in marked contrast to *n*MOSFETs, the measurement of 1/f noise does not represent a general degradation monitor for submicron *p* channel MOSFETs. Only at specific bias conditions is the 1/f noise sensitive to the device degradation. Secondly, the measurement of 1/f noise clearly provides a useful tool for investigating the various mechanisms of sub-micron *p*MOSFET device degradation. In particular, the insensitivity of the 1/f noise to device degradation at the condition $V_{gs} = V_{ds}$, supports the theory of hole trapping as the dominant degradation mechanism at this bias conditions. In addition, the 1/f noise results indicate slow state generation in the gate oxide for stress at $V_{gs} \approx V_{ds}/2$. Finally, from a technological perspective, the results are of relevance to the performance of *p*MOSFETs in analogue circuit applications. The results reveal that the precise operating point of the device will determine the long term stability of the *p*MOSFET 1/f noise.

## References

[1] J.M.Pimbley and G.Gildenblat, IEEE Elec. Device Lett, p.345 (1984)
[2] Z.H.Fang, S.Cristoloveanu and A.Chovet, IEEE Elec. Device Lett, p.371 (1986)
[3] C.Nguyen-Duc, Ghibaudo and F.Balestra., Proc. of ESSDERC, p.253 (1990)
[4] M.J.Deen and C.Quon, Proc. of INFOS, p.295 (1991)
[5] K.P.Cheung, S.Martin, D.Misra et al., Proc of IEDM, p.437 (1997)
[6] P.K.Hurley, E.Sheehan, S.Moran and A.Mathewson, Microelectron. Reliab., vol 36, p.1683 (1996)
[7] R.Woltjer, G.Paulzen, H.Pomp, H.Lifka and P.H.Woerlee, IEEE Trans. on Electron Devices, vol.42, p.109 (1995)
[8]. M. Stegherr, Solid State Electron., vol 27, p.1055 (1984).
[9] R.Jayaraman and C.Sodini, IEEE Trans. on Electron Devices, vol 36, p.1773 (1989).
[10] K.K.Ng and G.W.Taylor, IEEE Trans. on Electron Devices, vol 30, p.871 (1983)
[11] R.Woltijer, A.Hamada and E.Takeda, IEEE Trans. on Electron Devices, 40, p.392 (1993).
[12] T.Tsuchiya, K.Okazaki, M.Miyake and T.Kobayashi, Ext. Abstracts SSDM Conf., p.291, (1990)
[13] K.R.Hoffmann, Proc. of INFOS, p. 98 (1983)

PERGAMON

Microelectronics Reliability 38 (1998) 937–942

MICROELECTRONICS RELIABILITY

# Reliability of nitrided wet silicon dioxide thin films in WSi2 or TaSi2 polycide process : influence of the nitridation temperature

K. Yckache, P. Boivin, F. Baiget [a], S. Radjaa, G. Auriel, B. Sagnes, J. Oualid [b] and A. Glachant [c]

[a] *STMicroelectronics, ZI Rousset-B.P.2, F-13106 Rousset, France.*
[b] *LMCS EA 2192, ENSPM, Domaine Universitaire de Saint-Jérôme, 13397 Marseille cedex 20, France.*
[c] *CRMC2 (UPR-CNRS 7251), campus Luminy, case 913, F-13288 Marseille cedex 9, France.*

**Abstract**

In this paper it is investigated the influence of two polycide processes using WSi2 chemical vapor deposition (CVD) or TaSi2 deposition by sputtering on the integrity of wet and nitrided tunnel oxides ($t_{OX}$ = 6.6 nm) used in Electrical Erasable Programmable Read Only Memory (EEPROM) under negative Fowler-Nordheim electron injection (FNEI). It was confirmed that the nitridation reduces the generation rates of defects created by FNEI (interface states, neutral traps, positive and negative charges) and the stress induced leakage current (SILC).This reduction is more pronounced when $TaSi_2$ is used rather than $WSi_2$. In sputtering $TaSi_2$ process compared to the CVD $WSi_2$ one, it was observed : (*i*) - the generation of positive charges and the SILC are higher while the generation rates of interface states and negative charges are smaller ; (*ii*) - the generation rate of neutral traps is higher in wet oxide while it is smaller in nitrided oxides. Whatever the polycide process, the generation of the different types of defects created by FNEI is reduced by increasing the nitridation temperature while the SILC is increased. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Oxide reliability, specially resistance to Fowler-Nordheim electron injection (FNEI), is an important problem in scaling down Metal-Oxide-Semiconductor (MOS) devices and particularly in the reduction in endurance or charge retention of non-volatile memories, as Electrically Erasable Programmable Read Only Memory (EEPROM). It is well known that oxide integrity is improved by wet oxidation [1-2] and nitridation in $NH_3$, $N_2O$ or NO [3-5] when polysilicon gates are used. Since polycide gates are widely used in integrated circuit fabrication, it is important to study the influence of the silicide deposition on the oxide resistance to FNEI, specially on the integrity of tunnel oxide (thickness < 7 nm) in which electrons are injected to write or to erase the non-volatile memories. When a current is injected, the tunnel oxide requires [4] : (*i*) low charge trapping and consequently small defect generation rates to ensure a good endurance i.e. the highest number of write/erase cycles before degradation of the programming window; (*ii*) small leakage current to increase the charge retention in the floating gate and to reduce the consumption of electric energy. The continuous device scaling down makes these objectives more and more difficult.

In this paper, we compare the influence of $WSi_2$ or $TaSi_2$ deposition on the integrity of both wet and nitrided tunnel oxides. This integrity is defined by Stress Induced Leakage Current (SILC) on the one

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00120-6

hand and generation of defects by FNEI (interface states, charged and neutral bulk oxide traps) on the other, due to their contribution to the charge retention and the endurance degradation of non-volatile memories respectively.

## 2. Experimental

### *2.1 Technology*

The measurements were performed on simple Metal-Oxide-Semiconductor (MOS) capacitors (area A = 0.2 $mm^2$) fabricated without the metallization and passivation steps, on boron implanted substrate (effective doping $N_A = 5.10^{16}$ $cm^{-3}$). After field oxidation and sacrificial oxide, two different processes were used to grow the tunnel oxide. The wet oxide was grown in $O_2/H_2$ ambient at 770 °C during a time $t_{W1}$ followed by an annealing in nitrogen at 1000 °C during 10 minutes. The nitrided oxide was thermally grown in a conventional furnace in three steps : 1 - Initial oxidation in $O_2/H_2$ ambient at 770 °C during a time $t_{W2}$ ; 2 - Oxidation and nitridation in pure $N_2O$ at 900 °C, 950 °C or 1000 °C during times $t_{N900}$, $t_{N950}$, and $t_{N1000}$ ; 3 - Annealing in pure $N_2$ at 1000 °C for ten minutes. The times $t_{W1}$, $t_{W2}$, $t_{N900}$, $t_{N950}$, and $t_{N1000}$ were adjusted in order to obtain, for the different MOS capacitors studied, the same oxide thickness ($t_{OX} \approx 6,6$ nm) controlled by ellipsometry. The presence of nitrogen in the nitrided oxide was confirmed by Auger spectroscopy and by Secondary Ion Mass Spectroscometry. However, at this time, the dose was too low to be quantified with precision. We have observed, by Fourier Transform Infra Red analysis, that the incorporation of nitrogen is enhanced by the presence of OH radicals.

The $N^+$ polysilicon gate layer was deposited by the standard chemical vapor deposition process (CVD) and doped with Phosphorus. Then either a $WSi_2$ layer ($t_{WSi2}$ = 250 nm) was deposited by CVD from $SiH_4$ and $WF_6$ chemistry or a $TaSi_2$ layer ($t_{TaSi2}$ = 200 nm) was deposited by sputtering. It was previously confirmed that the $WSi_2$ process, which is generally used, induces a strong incorporation of fluorine in the oxide [6-7]. Finally, the different wafers were annealed at 420 °C in forming gas. Thus, height types of MOS capacitors were studied, labeled Wet/$WSi_2$, Wet/$TaSi_2$, N900/$WSi_2$, N900/$TaSi_2$, N950/$WSi_2$, N950/$TaSi_2$, N1000/$WSi_2$ and N1000/$TaSi_2$ respectively.

### *2.2 Characterization methods*

The different measurements were performed before and after two successive FNEI at constant current ($J_{FNEI}$ = - 16 mA/$cm^2$). The electrons were injected from the gate (gate voltage $V_G < 0$) at room temperature with injected charges $Q_{inj}$ = 0.16 C/$cm^2$ (FNEI-1) and 1.6 C/$cm^2$ (FNEI-2). During the constant current injection, the gate voltage shift ($\Delta V_G = V_G(t) - V_G(0)$) was recorded. The $\Delta V_G$ variations, during FNEI-1 and FNEI-2 respectively, are presented figures 1 and 2 for the Wet/$WSi_2$, Wet/$TaSi_2$, N900/$WSi_2$ and N900/$TaSi_2$ capacitors. As usually observed [8], $\Delta V_G$ is maximum for an injected charge $Q_{inj} \approx 0.1$ C/$cm^2$.

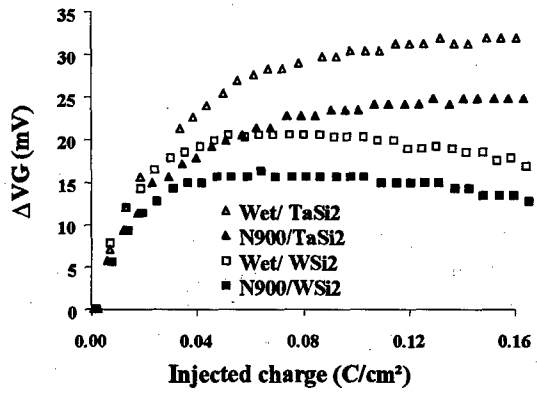



Figure 1 : Gate voltage shift during FNEI-1 ($Q_{inj}$ = 0.16 C/cm2, JFNEI = - 16 mA/$cm^2$)

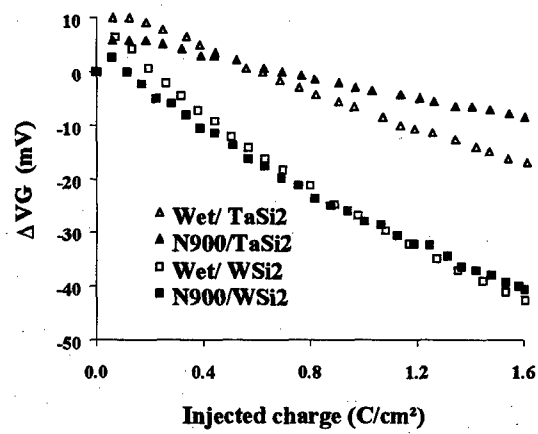



Figure 2 : Gate voltage shift during FNEI-2 ($Q_{inj}$ = 1.6 C/cm2, JFNEI = - 16 mA/$cm^2$)

It is shown in this figure, that the gate voltage shift is positive during the first injection and generally negative during the second injection

which indicates, as usual [3] [9], the creation of a net positive charge and a net negative charge respectively. The densities of the net positive and negative charges created during the first and second stress are directly related to $\Delta V_{Gmax}$ and to $\Delta V_{Gmax}$ - $\Delta V_{Gmin}$. $\Delta V_{Gmin}$ is the gate voltage shift at the end of the stress. It is important to point out that $\Delta V_G$ does not depend on the charges created at or near the oxide/substrat interface.

If $TaSi_2$ is deposited rather than $WSi_2$, it is clearly shown in Fig. 1 and 2 that : (*i*) the generation rate of positive charges is higher whereas the generation rate of negative charges is smaller; (*ii*) the beneficial influence of the nitridation on the generation rate of charges is more pronounced.

Before and after stress, the flat-band ($V_{FB}$), midgap ($V_{MG}$) and threshold ($V_T$) voltages were automatically measured from high frequency capacitance-voltage characteristics. The midgap voltage shift ($\Delta V_{MG}$,) depends essentially on the charges created in the oxide bulk near the substrate. The density of interface states ($N_{SS}$), created by FNEI, was determined from the difference of threshold voltage shift and flat-band voltage shift ($\Delta V_{FB}$ - $\Delta V_T$), assuming as usual that the interfaces states above and below midgap are acceptor-like and donor-like respectively. This measurements were confirmed by the results obtain from the classical quasi-static capacitance method. The results will be given in the next paragraph.

During FNEI, neutral electron traps are also created [10] which cannot be detected from $\Delta V_G$ or $\Delta V_{MG}$ measured just after stress. An avalanche electron injection (AEI) was performed with a fluence of $10^{18}$ e/cm$^2$ to fill the neutral electron traps created during the stress. AEI is obtained by biasing the gate with a sinusoidal voltage (frequency 100 KHz) to put the MOS capacitor alternatively in deep depletion and accumulation. During the deep depletion phase, hot electrons created by ionization are injected into the oxide. During the accumulation phase, electron-hole recombinations prevent the formation of the inversion layer. The AEI is periodically interrupted to measure the midgap voltage. The midgap voltage shift evolutions during AEI after FNEI-2 for the different MOS capacitors are presented in Fig. 3.

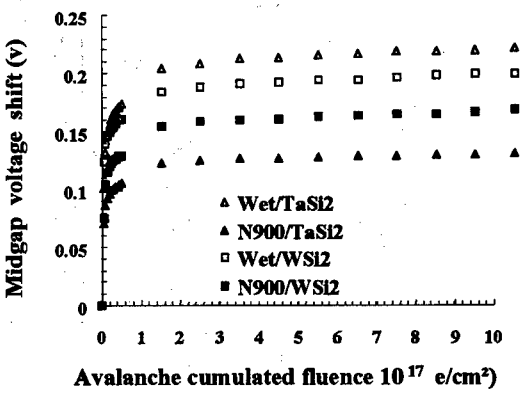



Figure 3 : Midgap voltage shift during Avalanche Electron Injection after FNEI-2 ($Q_{inj}$ = 1.6 C/cm$^2$and $J_{FNEI}$ = - 16 mA/cm$^2$)

These variations indicate that two different types of electron traps T1 and T2 were created by FNEI defined by their effective capture cross sections ($\sigma_1$ = 3 ± 1 $10^{-17}$cm$^2$, $\sigma_2$ = 2 ± 1 $10^{-18}$cm$^2$ respectively). The traps T1 are essentially responsible for the midgap voltage shift observed for low AEI fluence, up to 5 $10^{16}$ e/cm$^2$. The contribution of traps T2 is observed for AEI fluences higher than 5 $10^{16}$ e/cm$^2$. The same results were obtained on thicker oxides [11-13]. We have verified on fresh MOS capacitors that AEI did not modify the midgap voltage. Thus, the electron traps are effectively created by the FNEI applied prior the AEI. It is confirmed in Fig. 3 that the density of created electron traps is reduced by the nitridation. As for the positive and the negative charges, this reduction is larger when $TaSi_2$ is deposited rather than $WSi_2$. It can be observed that the generation rate of neutral electron traps in the Wet/$TaSi_2$ capacitors is slightly higher than in the Wet/$WSi_2$ ones when the contrary is true for the capacitors with nitrided oxide.

Current-voltage characteristics were performed before and after stress (Fig. 4 and 5) to measure the Stress Induced Leakage Current (SILC = $I_{after\ stress}$ - $I_{before\ stress}$). To eliminate the transient leakage current, the current-voltage characteristics were obtained using a negative staircase voltage ramp ($V_G$ < 0). The gate current was measured at each voltage step after a delay time (0.5 s).

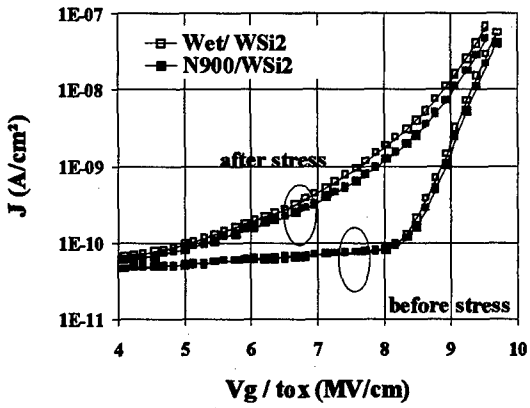



Figure 4 : Current-voltage characteristics before and after FNEI-2 ($Q_{inj}$ = 1.6 C/cm$^2$, $J_{FNEI}$ = - 16 mA/cm$^2$) for $WSi_2$ polycide process

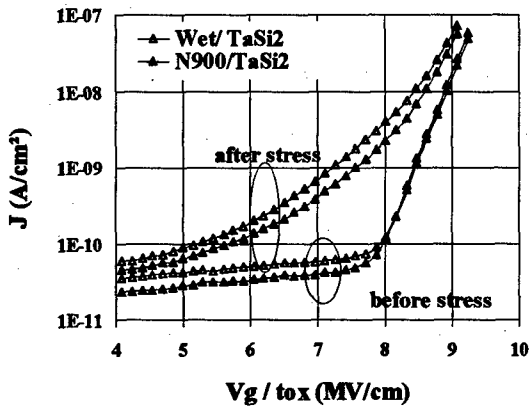



Figure 5 : Current-voltage characteristics before and after FNEI-2 ($Q_{inj}$ = 1.6 C/cm$^2$, $J_{FNEI}$ = - 16 mA/cm$^2$) for $TaSi_2$ polycide process

It can be observed that the leakage current measured before stress is not modified by the nitridation when $WSi_2$ is deposited. As for the generation of defects, in this case, the reduction of the leakage current after stress by the nitridation is smaller compared to the sputtered $TaSi_2$ polycide process. It can be noted, however, that the SILC is higher when $TaSi_2$ is used rather than $WSi_2$. These observations will be confirmed and discussed in the next section.

## 3. Results and discussion

In Fig. 6, 7, 8 and 9, we have reported respectively the mean values relative to the voltage gate shift $\Delta V_G$ during FNEI-1 and FNEI-2, to the densities of interface states ($N_{SS}$) and neutral traps ($N_{TI}$) created by FNEI-1 and FNEI-2 and to the SILC. These mean values were obtained from measurements performed on five MOS capacitors of each type. The dispersion of measurements was smaller than five percent.

In Fig. 6, it is confirmed that, during constant current FNEI from the gate, the gate voltage shift is positive if the injected charge $Q_{inj}$ is smaller than 0.16 C/cm$^2$ and is negative for $Q_{inj}$ = 1.6 C/cm$^2$. These variations are respectively related with the creation of a net positive and a net negative effective charge in the oxide layer. The positive charge results from the generation of trapped holes and ionized donor-like electron traps, while the negative charge is due to the creation of acceptor-like electron traps.

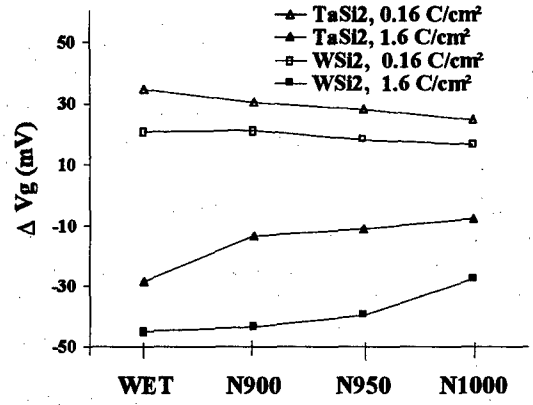



Figure 6 : Gate voltage shift at the end of FNEI-1 and FNEI-2

As generally observed [6] [7], the generation of positive and negative charges is reduced by the nitridation. This reduction is all the more important that the nitridation temperature is higher in accordance with previous results [4]. As already observed in the section 2, this reduction is more pronounced when $TaSi_2$ is deposited rather than $WSi_2$. It appears in particular that the nitridation has a very weak influence on the generation of positive charges when $WSi_2$ is used. It seems that the beneficial influence of the nitridation is partially masked by the $WSi_2$ deposition. It can be noted in Fig. 6 that the generation of positive charges is higher and the creation of negative charges is smaller in the case of the $TaSi_2$ polycide gate process compared to the case of the $WSi_2$ process. These observations indicate that the silicide deposition has a complex influence on the oxide resistance to FNEI.

The influence of silicide deposition is particularly noticeable on the generation of interface states as shown Fig. 7.

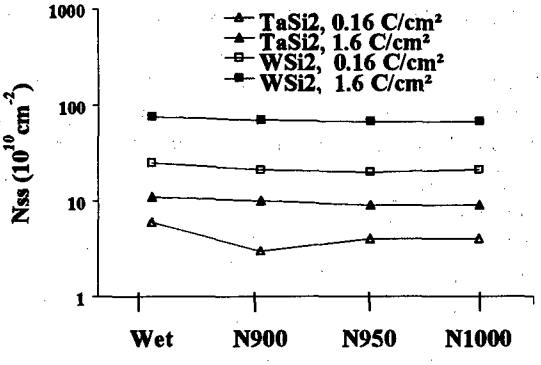



Figure 7 : Densities of interface states created by FNEI-1 and FNEI-2

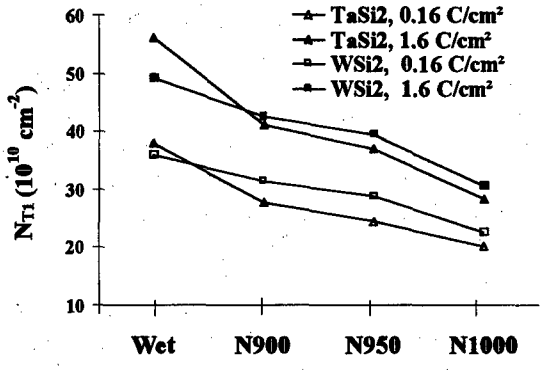



Figure 8 : Areal density of neutral electron traps $T_1$ created by FNEI-1 and FNEI-2

The generation rate of interface states is highly reduced when $TaSi_2$ is deposited in accordance with the charge-to-breakdown ($Q_{BD}$) values 3 C/cm$^2$ and 6 C/cm$^2$ for $WSi_2$ and $TaSi_2$ processes respectively measured using an exponential ramp current stress (ERCS) [1]. On the contrary, we have observed that the densities of interface states for virgin $TaSi_2$ capacitors is higher than for $WSi_2$ ones. This could indicate that some interface states are partially passivated during the WSi2 deposition, eventually due to the formation of Si-H bounds. During FNEI, this bounds could be broken and the hydrogen atoms which are liberated could contribute to the generation of new interface states. It can be noted that the nitridation has a weak influence on the generation rate of interface as well for the $WSi_2$ process than the TaSi2 one. This reduction of the interface states density due to the nitridation may be justified by the smaller amount of nitrogen atoms of nitrogen atoms observed after oxidation in a $N_2O$ ambient compared to NH3 or NO nitridation [3-5].

In the figure 8, it is clearly confirmed that the generation rate of neutral electron traps is all the more reduced as the nitridation temperature is increased. As observed for the generation rate of positive and negative charges, this reduction is more important when $TaSi_2$ is used rather than $WSi_2$ . In this case, the generation rate of neutral traps is higher in the wet oxide in accordance with the creation of positive charges when it is smaller in the nitrided oxide in relation with the creation of negative charges. These observations confirm the complex influence of the silicide deposition on the generation of the different types of defects created by FNEI. This influence depends on the oxidation process.

On the contrary, as shown in Fig. 9, the SILC measured for $V_G/t_{ox}$ = - 7 MV/cm is higher when $TaSi_2$ is deposited rather than $WSi_2$ as well as for wet than for nitrided oxide, excepted when the nitridation was performed at 1000 °C.

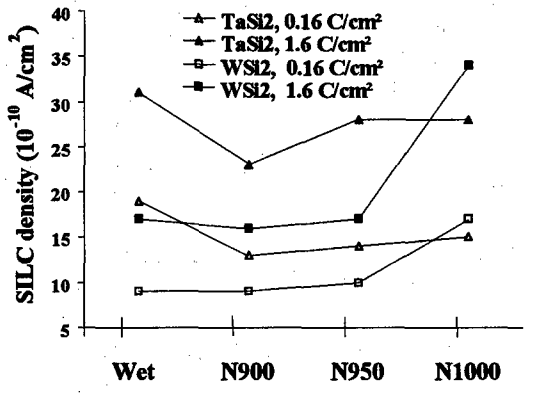



Figure 9 : Stress Induced Leakage Currents densities measured at Vg/tox = - 7 MV/cm$^2$after FNEI-1 and FNEI-2

However, it can be observed that the SILC is effectively reduced by nitridation when the TaSi2 is used in accordance with previous results [14]. But, it is not the case when $WSi_2$ is deposited. Contrarily to the generation of charges, neutral traps or interface states, the SILC increases with nitridation temperature. The origin of SILC in thin silicon dioxide films remains controversial. It was either attributed to the positive trapped charge generation [15-16] or to the neutral trap filling and emptying tunneling process [17-19]. The first hypothesis could explain the apparent correlation between the generation of positive charges near the injecting electrode (Fig. 6) and the SILC (Fig. 9). Effectively, the generation rate of positive charges and the SILC are higher for the $TaSi_2$ process than for the $WSi_2$

one. However, if the nitridation temperature is increased, we have observed that the generation rate of positive charges is smaller when the SILC is higher, both for the $TaSi_2$ and $WSi_2$ process. The second hypothesis seems to be more generally assumed but it appears that this hypothesis is not confirmed by the results reported in Fig. 8 and 9.

Thus, an apparent contradiction appears between, on the one hand, the increase of SILC and, on the other hand, the reduction of interface state, negative charge and neutral trap densities created by FNEI when $TaSi_2$ is deposited compared with $WSi_2$ deposition. Moreover, we have observed that the generation rates of the different defects are reduced when the SILC is increased by raising the temperature of nitridation. These contradictions could be explained if the SILC is assumed to be essentially dependent on localized oxide defects created by FNEI [20] while the measured generation rates of these defects are values averaged over the area of the MOS capacitors. The reduction of the defect generation rates and SILC by the nitridation is less pronounced when the $WSi_2$/polysilicon gate is used. It seems that the $WSi_2$ deposition partially masks the beneficial influence of nitrogen on the oxide resistance to electron injections. This observation could be associated with the high fluorine concentration measured in the oxide after $WSi_2$ deposition [6-7]. The role of incorporated fluorine in $SiO_2$ films are found to be quite complex [21]. Generally F-implanted oxides have been found to be more resistant to FNEI [22-23]. However, it was found that fluorine-less processes such as sputtered $WSi_2$ or $TaSi_2$ improve the integrity of nitrided oxides in reducing the generation of positive and negative charges and the creation of neutral traps compared to a fluorine-rich process such as $WSi_2$ deposition [6] [8]. Our results confirm these last observations. The effects of fluorine on oxide integrity are not yet understood and are particularly complex in the presence of nitrogen.

## 4. Conclusions

It was confirmed that nitridation of wet $SiO_2$ thin films reduces the SILC and the generation rates of interface states, positive and negative charges and neutral traps. This improvement of oxide integrity is partially masked when CVD-$WSi_2$ polycide process is used. This observation could be associated with the high fluorine concentration observed in the oxide after $WSi_2$ deposition. An apparent contradiction was observed between, from the one part, the generation rates of interface states, negative charges and neutral traps and, on the other, the generation rate of positive charges and the SILC which are respectively higher and smaller when $WSi_2$ deposition is used compared to the sputtered $TaSi_2$ process. This apparent contradiction could be justified if it is assumed that the SILC is associated to localized defects when the measured values of defect generation rates are averaged over the area of MOS capacitors.

## References

[1] E. Ciantar, P. Boivin, M. Burle, C. Niel, J.M. Moragues, B. Sagnes, R. Jerisian and J. Oualid, J. of Non-Crystalline Solids 187 (1995) 144-148.
[2] H. Sakake, N. Yasuda, S. Takagi and A. Tonumi, Appl. Phys. Lett. 69, 8, (1996) 1128-30.
[3] D. J. DiMaria and J.H. Stathis, J. Appl. Phys. 70 (1991) 1500-09.
[4] G. Ghidini, M. Alessandri, C. Clementi, D. Drera, and F. Pellizzer, J. Electrochem. Soc., vol. 144, n° 2 (1997) 758-764.
[5] L.K Han and D.L. Kwong, SPIE Austin vol. 2636 (1995) 190-197.
[6] Y. Shioya, S. Kawamura, I. Kobayashi, M. Maeda and K. Yanagida, J. Appl. Phys. 61 (1987) 5102-5109.
[7] J.M. Moragues, B. Sagnes, R. Jerisian, J. Oualid, E. Ciantar, J.L. Liotard and P. Merenda, Journal of Non-Crystalline Solids 187 (1995) 156-159.
[8] C.J. Chien, Y.C. Choi, G. Lo and C.D. Lien, SPIE Austin vol. 2636 (1995) 165-169.
[9] Y. Roh, L. Trombetta and J. Han, J. Electrochem. Soc. vol. 142, n°3 (1995) 1015-20.
[10] D.J. Dumin, S.K. Mopuri, S. Vanchinathan, R.S. Scott, S. Suramonian and T.G. Lewis, IEEE Trans Electron. Dev. 42 (1995) 760.
[11] M.V. Fischetti, J. Appl. Phys. 57 (1985) 2860.
[12] L.P. Trombetta, F.J. Feigh, R.J. Zeto, J. Appl. Phys. 69 (1991) 2512-21.
[13] G. Auriel, J. Oualid and D. Vuillaume, Microelectron. Reliab. 38 (1998) 227-31.
[14] L.K. Han, M. Bhat, D. Wristers, H.H. Wang and D.L. Kwong, Microelectronic Engineering 28 (1995) 89-96.
[15] Y.H. Lin, C.L. Lee and T.F. Lei, IEEE Trans. Electron. Dev. 45 (1998) 567.
[16] A. Scarpa, P. Ries, G. Ghibaudo, A. Paccagnella, G. Pananakakis, J. Brini, G. Ghidini and C. Papadas, ESSDERC Stuttgart (1997) 592-595.
[17] R. Rofan and C. Hu, IEEE Electron. Dev. Lett. 12 (1991) 632.
[18] D.J. Dumin and J.R. Maddux, IEEE Trans. Electron. Dev. 40 (1993) 986.
[19] D. J. DiMaria and E. Cartier, J. Appl. Phys 78 (1995) 3883-94.
[20] P. Olivo, T.N. Nguyen and B. Ricco, IEEE Trans. Electron. Dev. 35 (1988) 2259.
[21] T.K. Nguyen, L.M. Landsberger, C. Jean and V. Logiudice, IEEE Trans. Electron. Dev. 44 (1997) 1432-40.
[22] Y. Nishioka, K. Ohyu, Y. Ohji, N. Natuaki, K. Mukai and T.P. Ma, IEEE Electron. Dev. Lett. 10 (1989) 141-143.
[23] P.J. Wright, N. Kasai, S. Inoue and K.C. Saraswat, IEEE Electron. Dev. Lett. 10 (1989) 347-348.

PERGAMON

Microelectronics Reliability 38 (1998) 943–950

MICROELECTRONICS RELIABILITY

# Assembly and analysis of quantum devices using SPM based methods

Lars Montelius, T. Junno, S.-B. Carlsson, M. H. Magnusson, K. Deppert, H. Xu, and L. Samuelson.

*Division of Solid State Physics &The Nanometer Structure Consortium, Lund University, P.O. Box 118, S-221 00 Lund, Sweden*

**Abstract**

Presently, there is a rapid growing interest in quantum devices based on single-electron-tunneling (SET)-effects. Such SET-devices are possible candidates for single electron memories that might allow room temperature operation. The focus of possible room temperature SET-devices is towards fabrication of lateral tunnel structures. We have during the last years developed a new way to fabricate lateral quantum devices. The principle is based on the use of the atomic force microscope (AFM) for very accurate movement of pre-fabricated nano-objects into a small gap between two metal electrodes. In this way a unique technique to build extremely small scale and accurate structures for contacting of nano-objects is realized, making hitherto unachievable devices to be formed. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

The dimensions of electronic devices have been reduced approximately a factor of two every three years for the last 20 years, effectively meaning an increase of capacity with a factor of four. This trend, known as Moore´s law, is predicted to hold for the next decade as well, although small deviations are foreseen [1,2]. This means that around year 2010 the linewidths of the components produced in large scale will be below ≈100 nm. This presents a great challenge not only to the technology community for fabrication of these devices, but also to the electronic design people since the electronic properties of nm-scale devices will be governed by quantum mechanics. Therefore, new paradigms for realization of logical devices might appear in the near future. One such nanostructure paradigm is the logic based on "Quantum Cellular Automata", originally proposed by Lent et. al. [3,4] and recently demonstrated. although at low temperatures, in a paper by Orlov et. al. [5].

As device sizes are shrinked to -and below- the nanometer size range it is also possible that molecular materials will come into play, especially in combination with "ordinary" electronic materials. Already today several such papers have been published where molecules have been employed as functional elements in different kinds of devices [6].

In this paper will be discussed different concepts for realization of nanoscale devices, with preference for structures fabricated using scanning probe based methods. The different nanodevices that will be discussed and introduced are i) nanomechanical switches and components whose performance are governed by ii) single electron tunneling effects and iii) quantized conductance phenomena. For realization of these classes of

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00165-6

components we have since some years back been developing an AFM-based method where prefabricated nanostructures are assembled into functional units. This technique will first of all be described in the next paragraph "Experimental details", and then each of the above classes of components will be introduced and discussed in separate paragraphs.

## 2. Experimental details

The principle is based on the use of the atomic force microscope (AFM) for very accurate movement of pre-fabricated nano-objects. These nano-objects can be of various kinds, such as lift-off defined metal discs, aerosol fabricated nano-particles [8,9] and colloidal gold particles. The assembly procedure for making the various components discussed above, is described in detail in a paper by Junno et. al. [7], but consists primarily of moving selected nanometer sized objects into a small gap between two metal electrodes (see Fig 1).

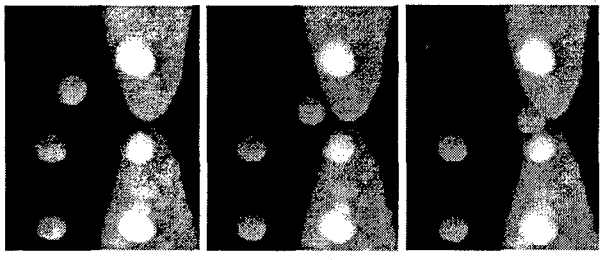


Fig. 1. Three AFM images (740 nm x 740 nm) recorded during the manipulation of an Au disc into a gap between two Au electrode.

The metal electrodes having gaps between 10 nm and ≈100 nm consists of a 30 nm Au-layer on top of a 3nm Ti-layer. The electrodes are defined by conventional e-beam lithography in combination with lift-off technique. The Ti layer acts as an efficient adhesion agent for the subsequent Au-layer. The Si substrate employed has a 300 nm thick $SiO_2$ top layer grown by thermal oxidation on its surface. The e-beam lithography is made with a commercial scanning electron microscope, (SEM), equipped with a $LaB_6$ filament for achieving a stable enough filament current suitable for lithography applications. The SEM has been further modified into an e-beam writer by using software and hardware from Raith GmBh [10].

The method [7] for imaging and manipulation can be described as follows: i) make an image in non-contact mode, ii) select the particle to be moved, iii) position the tip slightly away from the particle, iv) open the feed-back loop, v) advance the tip into contact with the particle, vi) push the particle by the tip induced lateral forces to the desired position, vii) withdraw the tip, viii) establish feedback and acquire a new image displaying the outcome of the particle translation. This procedure, (i-viii), is repeated until the particle is moved to the desired location. The rate when pushing the particles is approximately 100 nm/s, and the particles can be pushed (or moved) with nanometer precision over large distances, i.e. several hundreds of nanometers, if needed.

The AFM-based imaging and manipulation is done in ambient air condition at room-temperature using an ultra-sharp non-contact AFM tip. For imaging we used an oscillation frequency of around 170kHz together with conventional lock-in detection of the amplitude change. The atomic force microscope employed is commercially available [11], although the manipulation procedure is developed in-house by writing subroutines to the scanning program supplied with the microscope.

A prerequisite for moving nanosized particles is the sharpness of the tip employed. If the tip is too dull, then particles smaller than the tip radius have a tendency to adhere to the tip when the tip is withdrawn after the pushing action. This can be understood easily using a simple model where the last part of the tip is approximated by a sphere with a tip radius $R_t$, and assuming that the particle to be moved is sperical with a radius $R_p$. Since we are operating in ambient air and at room temperature, the interactions between particle-tip and particle-substrate respectively, will have contributions from both the van der Waals forces and the capillary condensation forces. The attractive adhesion force between particle and substrate, $F_{p\text{-}s}$, can be written as [12]:

$$F_{p\text{-}s} = 4\pi R_p \gamma_{sv} \qquad (1)$$

where $\gamma_{sv}$ denotes the interfacial energy of the solid-vapor interface. Then the attractive force

between the particle and the tip, $F_{p-t}$, can be written as (by replacing $R_p$ in eq.1 with $R_tR_p/(R_t+R_p)$) :

$$F_{p-t}= 4\pi (R_tR_p/(R_t+R_p)) \gamma_{sv} \quad (2)$$

Eq. 2 shows us that we have to minimize the tip radius $R_t$ (assuming $\gamma_{sv}$ to be equal for the particle-tip and particle-substrate interactions), in order to avoid that the particle get stuck to the tip instead of remain at the surface of the substrate, i.e. to ensure that $F_{p-s}> F_{p-t}$. In particular, if $R_t=R_p$ then $F_{p-t}$ is only half the value of $F_{p-s}$.

The commercial relatively sharp AFM tips have been further processed by EBD (E-beam deposition), resulting in a carbon-rich ultra sharp tip [see Fig. 2).

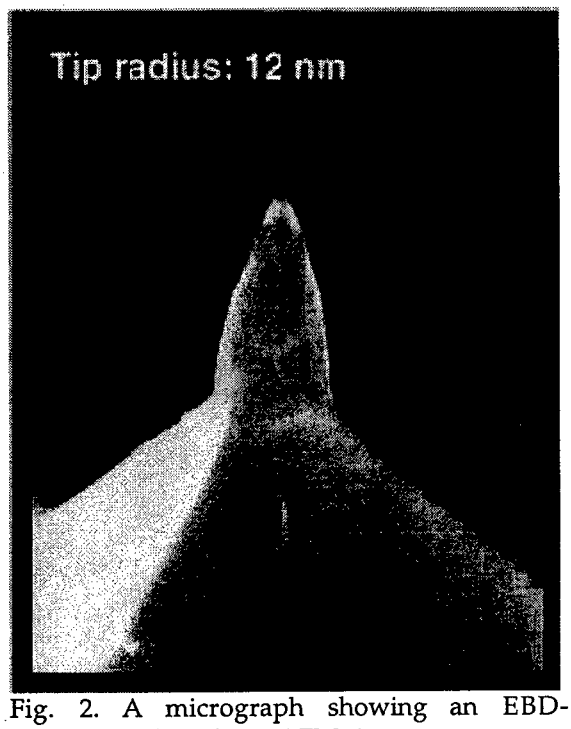


Fig. 2. A micrograph showing an EBD-processed ultra sharp AFM tip.

The EBD is performed by hitting the tip, mounted into a field-emission scanning electron microscope, (FE-SEM), with the focused electron beam onto the very tip apex. Thereby, the always present hydrocarbon fumes originating from the FE-SEM pump-oils, are cracked resulting in a sharp carbon-rich supertip. We have grown tips having lengths between 0.5μm-1μm with tip radius of 15-25 nm. Sometimes, the EBD tips have been further sharpened by putting the tips into an oxygen plasma, thereby a typical tip radius in the range of 10-20 nm can be obtained.

The AFM we work with scans the tip instead of the sample. This scan principle has made it possible to modify the sample holder stage, permitting mounted and bonded chips onto chip carriers (DIL) to be used [15]. Since the tip is scanned and the sample is not moved during neither the scanning nor the manipulation, electrical in-situ monitoring during the actual assembly procedure is possible [13,14,15]. Furthermore, when a certain structure is fabricated using the AFM-based method and tested initially in-situ, it can easily be transferred to other measurement set-ups, such as e.g. cryostats enabling low-temperature measurements to be performed.

One of the major benefit with our method is the possibility to electrically monitor the whole assembly procedure by having a potential applied between the electrodes. This means that electrical feed-back control of the fabrication of a quantum device is possible, allowing fine tuning of the wanted electrical properties of a specific component.

## 3. Results and discussion

The techniques of pre-fabricating nm-scale building blocks and AFM-manipulation with simultaneous electrical measurements have allowed us to build nano-mechanical switches and nano-bridges, quantum transport devices and single-electron lateral devices with control on the Ångström level [13,14,15]. This assembly procedure of manipulating small pre-fabricated objects into a prefabricated electrode structure in a nano-LEGO fashion (see Fig. 2 and 3), is to be seen as a versatile tool for nano-scale contacting and investigation of nano-objects.

### *3.1 Nanomechanical switches and bridges*

In this case the electrode gap was adapted to be of a size that allows one single Au-disc to bridge the gap. The Au-discs employed were fabricated after the electrode structure was made in a second e-beam lithography and lift-off step.

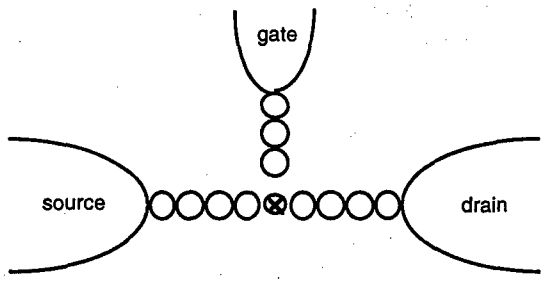



Fig. 3. A principal sketch showing the electrode layout employed for making the different nanodevices.

The Au-discs were 30 nm thick and had diameters between 30-100 nm, and were defined in a grid on the sample surface. The electrical measurements were performed using either a 2- or 4-terminal scheme having a constant applied bias in the range of 1-100mV. When making the ohmic connection firm contacts is established between the metal disc and both of the electrodes. Then, the resistance value between the electrodes is typically 120 ohm, which is the same value as if one would have closed the gap with e-beam lithography directly in the first e-beam lithography step. Thus, no extra resistance drop is found due to the particle electrode junction, opening up the possibility to use this procedure for fabricating nanobridges between particles and featues on surfaces. Furthermore, by moving the metal disc in and out of contact with one of the electrodes, the resistance value will toggle between these 120 ohms in the "on-state" to values in the TOhm regime in the "off-state" (see Fig. 4), just as a bi-stable nanomechanical switch [15]. During this experiment the on-state current level was reproducible to within 1%.

### *3.2. Single electron tunneling components*

Presently, there is a rapid growing interest in quantum devices based on single-electron-tunneling (SET)-effects. Numerous metal and semiconductor based tunnel barrier structures have been shown to exhibit Coulomb blockade effects at low temperatures. The Coulomb blockade effect is manifested if the charging energy, $E_c$, of the charging element utilized for the Coulomb blockade effect, is larger than the thermal energy, $E_t$.

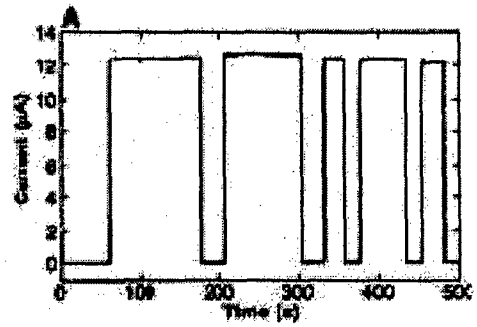



Fig. 4. The figure shows the current through the "nanomechanical bridge" as a 50 nm Au particle is pushed into and out of contact with two Au/Ti electrodes.

The charging energy is given by

$$E_c = q^2/2C \quad (3)$$

where q equals the elementary unit of charge and C equals the capacitance of the charging island given by:

$$C = epsA/d \quad (4)$$

where eps=dielectric constant and A=area of the island ($A=\pi r^2$ if the island is spherical with radius r) and d=thickness of the tunnel barrier. The thermal energy is given as usual by the expression:

$$E_T = kT \quad (5)$$

where k=Boltzmans constant and T=temperature, which gives a value of $E_T \approx 26$ meV at room temperature.

A normally employed safety marginal requires the charging energy to be more than 20-30 times larger than the thermal energy (or noise), which means that the size of the charging island must be around some few nanometers for SET-devices to function at room temperature [16,17]. Indeed this was accomplished by Scönenberger et. al. [18] using the tip of a scanning tunneling microscope and a small metal particle, creating a vertical double barrier structures where observation of Coulomb charging effects at room temperature was observed [18]. However, from a practical point of view it is obvious that scanning tunneling microscopes cannot

form the basis for any electronic application where single or complex combinations of SET devices will be employed.

Lately the focus of possible room temperature SET-devices have instead been directed towards fabrication of lateral tunnel structures. For instance, Chen et. al. [19] deposited AuPd nanocrystals in-between Au electrodes defined by e-beam lithography, and they reported clear Coulomb blockade effects at 77K. However, the fabrication method is very hard to control and reproduce, which limits it's practical use for making devices. Other techniques that might have a larger potential for device fabrication involves some kind of self-aligning/assembly process(es) of conducting particles between metal electrodes, as reported for instance by Klein et. al. [20]. They managed to make a link of several colloidal Au clusters between e-beam defined electrodes. The electrodes were chemically modified in order to achieve good adhesion of the colloidal particles. However, large fluctuations in the obtained I-V spectra due to environmental changes during the measurement makes also this method not optimal. A very similar approach was reported by Sato et. al. recently [21], where a chain of colloidal particles made up a quasi one-dimensional channel between two Au-electrodes. The device exhibited clear Coulomb blockade effects at both 4.2 and 77K.

In our scheme, we have full control of the fabrication procedure due to the electrical in-situ monitoring possibility. Since the microscope is operated at room temperature we usually assemble the structure firstly, and then transfer the complete device into a cryostat allowing variation of the temperature during the measurements. The building blocks that we have used for realization of SET structures are primarily based on aerosol defined clusters having sizes around some few 10's of nanometers. In fig. 6 below is shown the outcome of using three 30 nm In aerosol particles manipulated into a less than 100 nm electrode gap region. The In particles are coated with a native oxide functioning as the tunnel barrier. As shown in figure 4 a clear Coulomb blockade region is obtained at 4K. In the figure is also shown the electrical characteristics at room temperature obtained during the actual assembly of the nanodevice. The Coulomb blockade effect is of course thermally smeared out at room temperature since the In particles are too large for observing SET-effects at room temperature. Recently, we have obtained Coulomb oscillations in similar structures by employing a gate voltage and thereby unambiguously controlling the charge on the central island.

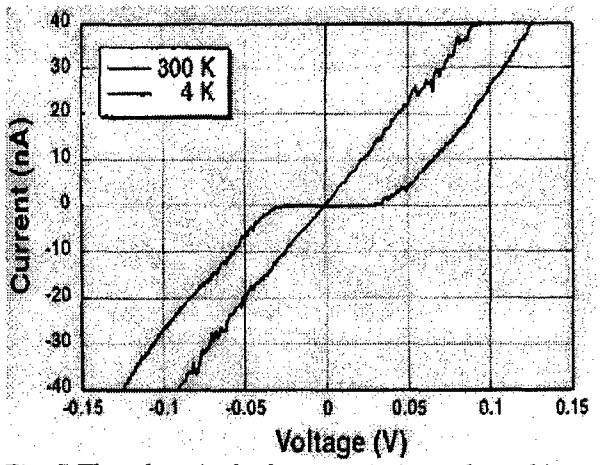



Fig. 5 The electrical characteristics when the Au/Ti electrodes are bridged with three In 30 nm aerosol particles. At 300K the I/V relation is linear and at 4K a clear Coulomb blockade effect is displayed.

### *3.3 Quantized conductance components*

For narrow constrictions of the order of the Fermi wavelength of the electrons in the material, the resistance will be quantized in units of the resistance quantum, $R_0$, given only by a relation between the fundamental constants q and h, Planck's constant [22]. If the experimental conditions are fulfilled then the conductance of such a narrow constriction will be quantized showing up as a ladder with steps of an integer number of the conductance quantum, $G_0=2q^2/h$. The experimental requirements are most easily fulfilled for semiconductors since then the Fermi wavelengths are in the order of ≈300Å, as compared to a metallic system where the Fermi wavelenght is typically ≈5Å. But since the energy level spacing of a typical semiconductor is around 0.1 meV, then the temperature has to be lowered to the mK range for observation to be possible. The first examples [23] of quantized conductance was measured in a GaAs/AlGaAs 2-D

electron gas, having a top split gate geometry that defines a narrow constriction point, a quantum point contact (QPC). The current measured between source and drain electrodes contacting the 2-D electron gas, exhibited a plateau like staircase conductance.

Room temperature measurements of such phenomena requires a metallic system to be employed, since then the energy level spacing is in the order of ≈1 eV>> the room temperature thermal energy. Thus the difficult requirement to cope with is the forming of a metallic narrow constriction in the range of some few Å. This has been possible using a so-called mechanical break-junction [24-26) in which a metallic constriction is positioned on top of a piezoelectric actuator that will make the metallic constriction to bend and very slowly deform, until a break junction is formed. By measuring the current through such a break junction is possible to observe quantized conductance. However, the hysteresis is quite large both due to mechanical vibrations and piezoelectric creep etc, making the measurements rather complicated and time consuming.

Recently. a popular method for room temperature measurements has been to employ an STM tip and hit the tip into a metallic surface. The intendation of the tip into a metallic sample surface and the subsequent withdrawal from the sample surface creates a range of quantized conductance channels [27-30]. The process of forming the tiny metallic wires can be described as very similar to Swiss fondue. By simultaneously measure the current between the tip and the sample it is possible to detect the quntized conductance by adding up histograms of several thousands of such intendations of the tip into the sample surface. Indeed, this also works fine with macroscopical objects [31], for instance by measuring the current between the electrodes in a mechanical relay as a function of time is straight forward to visualize quantized conductance.

However, all of the methods discussed above with regards to room temperature measurements has drawbacks, since it is impossible to form a certain size of the narrow constriction that is time invariable. Using our scheme of manipulating matter with the tip of the AFM while monitoring the conductance by having a potentail applied between the elctrodes, it is possible to form such stable constrictions [14,15]. The potential applied varied between 1-100 mV and the current was recorded continously in a 2-terminal measurement scheme. As an example is shown in Fig. 6 the experiment in which a quantized point contact (QPC) is formed when a metal disc of Au is pushed out from the two metal electrodes after having been in firm contact.

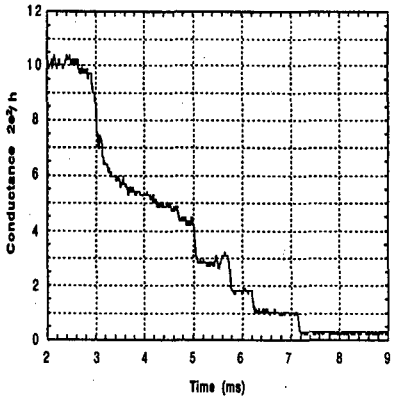



Fig. 6. Dynamic behavior of the conductance as an Au nanodisc is being pushed out of contact with the electrodes. The applied bias voltage was 2 mV and the temperature 300 K.

During this operation it is secured that the disc is in firm contact only at one electrode, then at the other side a very weak mechanical contact is formed (see Fig. 7). The same characteristics are obtained when a particle is pushed slowly into contact.

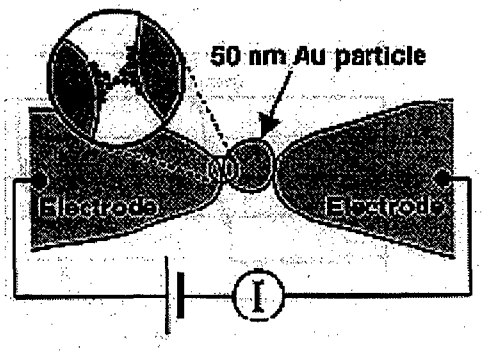



Fig. 7. Visualization of how tiny metallic necks are formed when a particle is pushed very slowly out/in between the electrode gap.

As a function of how far one push this side of the metal disc out of contact with the electrode, a quantized conductance staircase is obtained, having plateaus in the units of $2e^2/h$. Thus, the nano-junctions formed between the disc and the electrode by these gentle pushes are in the range of some few Ångströms, i.e. atomic dimensions.

Furthermore, by pushing the particles just small steps, nominally 1Å per push, it is possible to stabilize the current at a certain conductance plateau level for timescales up to several 10's of minutes. In this condition, the current is stable and decoupled from environmental fluctuations since the mass of the particle is so tiny that gravitational forces are not in affect (see Fig. 8).

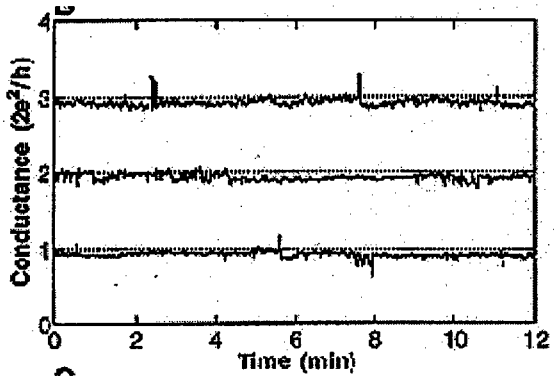



Fig. 8. Typical time stable conductance plateaus obtained by pushing a 50 nm Au disc in or out of contact with one of the electrodes, thereby forming connections on the Å-scale.

Eventually, the junction breaks. The reason for this is unclear but might be due to electromigration effects or nanoparticle dynamics. There should not be any losses within the QPC region since it is ballistic, but close to this region the current densities might nevertheless be high enough to cause electromigration to occur.

## 4. Outlook

This method of assembling structures using predefined nanosized object is to be seen as a versatile tool for nano-scale contacting and investigation of both inorganic and organic nano-objects, e.g. a particle covered with an organic "skin" (see Fig. 4 above). In this way it is possible to study the transport of carriers from an electrode to a molecule and then to the particle or chain of particles. The fields of biophysics and "molectronics" (molecular electronics) are expected to benefit of this general technique, both for basic studies and fore more applied studies.

The usage of SET-effects may also in the near future come to play an important role for diagnostics of devices and components, since it is possible to implement a SET-structure in the very end of an SPM-tip. Thus, if the dimensions of the charging island employed for the SET effect is small enough, a room temperature highly efficient and very sensitive (singe charges) charge probe with a very high spatial resolution could be a reality.

## Acknowledgment

This work has been performed within the "Nanometer Structure Consortium" in Lund, and we gratefully acknowledge support from the Swedish National Board of Technical and Industrial Development, the Swedish Natural Science Research Council, the Swedish Technical Research Council, The Foundation of Strategic Research, the Crafoord Foundation and ESPRIT proj no: 22953.

## References

[1] Road-map of Semiconductor Industry Association (SEMATECH, Austin Texas)

[2] Road-map of Microelectronics-Advanced Research Initiative (MEL-ARI, EU, Brussels)

[3] C. S. Lent et.al., Nanotechnology **4**, 49 (1993)

[4 ] C. C. Lent and P. D. Tougaw, Proceedings of the IEEE **85**, 541 (1997)

[5] A. O. Orlov et. al. Science **277** , 928 (1997)

[6] See for instance: M. A. Reed et al., Science **278**, 252 (1997)

[7] T. Junno, K. Deppert, L.Montelius and L. Samuelson, Appl. Phys. Lett. **66**, 3627 (1995)

[8] K. Deppert, et al., J. Crystal Growth **169**, 13 (1996)

[9] K. Deppert and L. Samuelson, Appl. Phys. Lett. **68**, 1409 (1996)

[10] Raith GmBh, Germany
[11] Explorer 1000, Topometrix, Santa Clara, USA
[12] J. N. Israelachvili, Intermolecular and surface forces, (Academic Press, New York, 1985
[13] L. Montelius, T. Junno, S.-B. Carlsson and L. Samuelson, Microel. Engin. **35**, 281 (1997)
[14] T. Junno, S.-B. Carlsson, H. Xu, L. Montelius and L. Samuelson, APL **72**, 548 (1998)
[15] S.-B. Carlsson, K. Deppert, T. Junno, M. H. Magnusson, L. Montelius and L. Samuelson, Semicond. Sci. Technol. **13**, 1 (1998)
[16] D. V. Averin and K. K. Likharev, J. Low. Temp. Phys. **62** 345 (1986)
[17] T. A. Fulton and G. J. Dolan, Phys. Rev. Lett. **59**, 109 (1987)
[18] C Schönenberger, H. van Houten and H. C. Donkersloot, Europhys. Lett. **20**, 249 (1992)
[19] W. Chen, H. Ahmed and K. Nakazato, Appl. Phys. Lett. **66**, 3383 (1995)
[20] D. L. Klein, P. L. McEuen, J. E. Bowen Katari, R. Roth and A. P. Alivisatos, Appl.Phys. Lett. **68**, 2574 (1996)
[21] T. Sato, H. Ahmed, Appl. Phys. Lett. **70**, (1997)
[22] R. Landauer, J. Phys.: Condens. Matter **1**, 8099 (1989)
[23] B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. van der Marel and C. T. Foxon, Phys. Rev. Lett. **69**, 140 (1988)
[24] C. J. Muller, J. M. v. Ruitenbeck, and L. J. d. Jongh, Phys. Rev. Lett. 69, **140** (1992)
[25] J. M. Krans, C. J. Muller, I. K. Yanson, R. H. Th. C. M. Govaert and J. M. v. Ruitenbeck, Phys. Rev. B **48**, 14721 (1993)
[26] J. M. Krans, J. M. v. Ruitenbeck, V. V. Fisun, I. K. Yanson and L. J. d. Jongh, Nature **375**, 767 (1995)
[27] N. Agrait, J. g. Rodrigo, and S. Veira, Phys. Rev. B **47**, 12345 (1993)
[28] J. I . Pascual, J. Mendez, J. Gomez-Herrero, A. M. Baro, N. Garcia, U. Landman, W. D. Luedtke, E. N. Bogachek and H.-P. Cheng, Science **267**, 1793 (1995)
[29] J. I . Pascual, J. Mendez, J. Gomez-Herrero, A. M. Baro, N. Garcia, Phys. Rev. Lett. **71**, 1852 (1993)
[30] K. Hansen, E Laegsgaard, I. Stensgaard, F. Besenbacher, Phys. Rev. B **56**, 2208 (1997)
[31] J.-L. Costa-Krämer, N. Garcia, P. Garcia-Mochales and A. P. Serena, Surf. Sci. Lett. **342**, L1144 (1995)

PERGAMON

Microelectronics Reliability 38 (1998) 951–956

MICROELECTRONICS RELIABILITY

# Circuit internal logic analysis with Electric Force Microscope- (EFM-) testing

J. Bangert and E. Kubalek

*Werkstoffe der Elektrotechnik, Gerhard-Mercator-Universität Duisburg, Bismarckstr. 81, 47048 Duisburg, Germany*

Abstract

In this paper the well known electron beam testing techniques stroboscopic voltage contrast and logic state mapping are realized for an EFM-testing system. These testing techniques allow an advanced logic analysis of digital circuits. They are based on a linear interaction, a pulse sampled measurement method and a scanned EFM-tip. First measurement results are shown on an integrated bus structure of 2 μm feature size with applied clock signals at few megahertz frequencies pointing out the ability for simultaneous temporal and spatial frequency measurements with EFM-testing, as well. Further improvements promise the applicability of these proven testing techniques in integrated circuits with structure dimensions beyond the limits of electron beam testers. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

The ability to perform non-invasive logic analysis at internal test points of integrated circuits (IC) is indispensable in the design and test of microelectronics. Voltage coding, stroboscopic voltage contrast and logic state mapping are well known and suitable testing techniques because they allow a logic analysis and an orientation inside of the IC simultaneously. These measurement techniques are available for the Electron Beam Tester (EBT) for logic analysis and inspection [1]. But practical problems and the almost achieved theoretical margins of the EBT were reported. Especially the required high spatial resolution for the analysis of nowadays ICs is problematic. A practical limit of 0.7 μm is reported [2]. The measurement errors for 0.5 μm wide conducting lines are in the range of 20 %.

Electric force microscopy- (EFM-) testing has been reported as a new testing system for IC internal contactless diagnosis without any IC preparation and without restrictions in respect to ambient conditions. High spatial resolution in the sub-micrometer range and millivolt sensitivity were achieved simultaneously for periodic dynamic signals [3,4]. A spatial resolution of 80 nm was shown at frequencies up to 10 GHz [5]. Qualitative measurements of digital patterns were done several times at distinct test points at GHz-clock rates [4,6,7], quantitative amplitude measurements of digital signals with MHz-clock rates were also shown [8,9]. Though these results prove that EFM-testing is very promising there is still a lack of testing techniques for logic analysis. This paper reports on the realization and application of the testing techniques stroboscopic voltage contrast and logic state

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00072-9

mapping with an EFM-testing system. The underlying ideas and basic simulations of the interaction force are presented. An experimental set-up for these testing techniques is realized and incorporated in an existing, in-house developed EFM-testing system. The present state of the performance is demonstrated by means of dynamic voltage measurements at 2 MBit/s on a bus structure with 2 µm feature sizes within the ITG test chip. The good chances achieving a better spatial and temporal resolution will be evaluated and discussed.

## 2. Principle of operation and measurement set-up

The electric force microscope- (EFM-) testing is based on the Coulomb force interaction between the EFM-probe and the test point located on a conducting line at $(x,y)$ of the device under test (DUT). The EFM-probe consists of a conducting miniaturized and sharp measurement tip, positioned below a conducting cantilever. The measurement tip height is about 15 µm (see Fig. 1a)) The cantilever has a length of 450 µm and is 60 µm wide. Based on a simple plate capacitor model of the measurement tip, ignoring the cantilever, and a small area of the conducting line below the measurement tip, the measurement principle leads towards a quadratic formula [3]

$$F_z(x,y) \sim \left(u_p(t) - u_m(t,x,y)\right)^2 = u_p^2(t) + u_m^2(t,x,y) - 2 \cdot u_p(t) \cdot u_m(t,x,y) \sim \Delta h \quad (1)$$

where $F_z(x,y)$ is the Coulomb force normal to the conducting line surface, $u_p(t)$ the external controlled time variant voltage on the EFM-probe, $u_m(t,x,y)$ the unknown voltage at the test point on the conducting line of the DUT and $\Delta h$ the cantilever bending caused by the force $F_z$. Experiments and calculations showed that the cantilever-sample capacitance has to be taken into account for the force calculation and the cantilever bending [10,11].

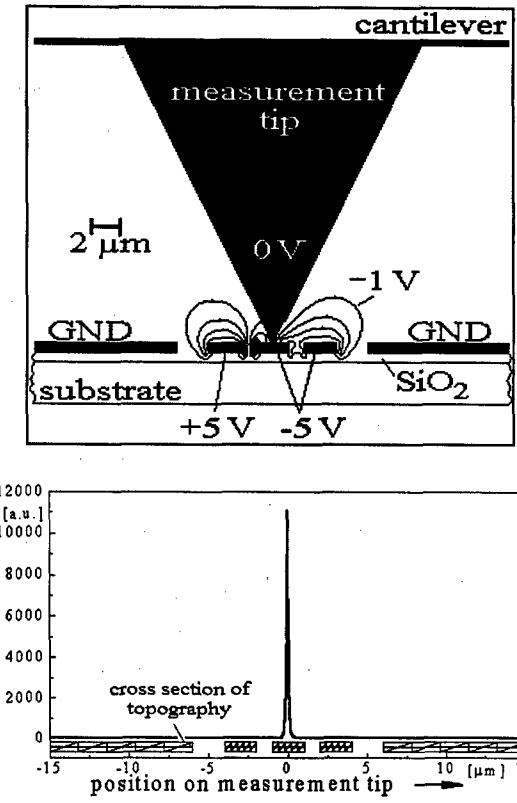



Fig. 1. Simulations of the force distribution along the measurement tip above a three conducting line bus structure: a) equi potential lines for a measurement tip centred above the middle conducting line, every equi potential line represents a difference of 1 V b) force distribution of interaction forces

A new type of simulations of the potential and the force distribution is based on a static two dimensional finite difference method [12,13]. The potential distribution is calculated from a boundary value problem and the force distribution from the electric fields. This is an advanced approach in comparison to the plate capacitor model. The shapes of the cantilever, the measurement tip and a bus structure are depicted in figure 1 and figure 2. The bus structure consists of three 2 µm wide conducting lines with 1 µm spacing surrounded by ground planes and a height of 0.5 µm on a insulating layer. The chosen potentials are 0 V for the ground potential (GND) and the measurement tip, +5 V for the left and -5 V for the middle and the right conducting lines.

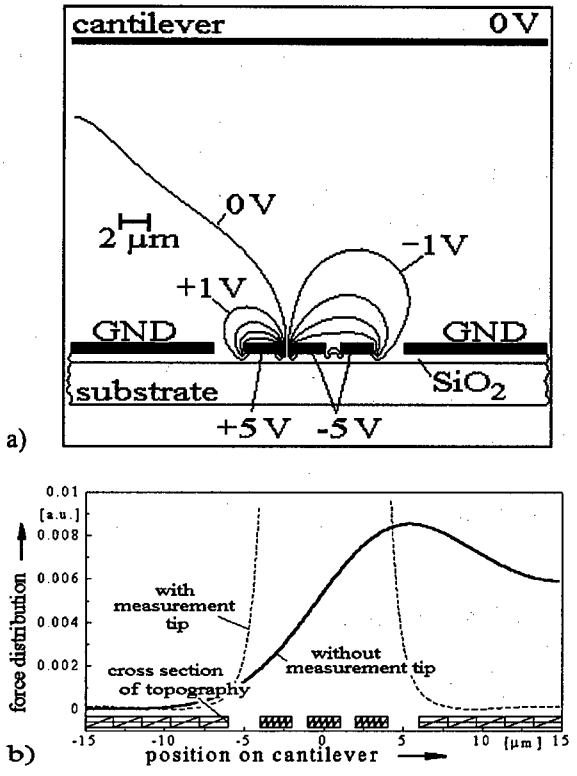



Fig. 2. Simulations of the force distribution along the cantilever above the bus structure: a) equi potential lines for a cantilever in a distance of 15 µm b) comparison of the force distribution with and without measurement tip (identical $y$-scale for both plots)

The potential difference between the measurement tip and the bus structure causes the interaction forces. Figure 1 depicts the geometries, the equi potential lines and the force distribution along the measurement tip shape. The forces are concentrated around the very end of the measurement tip. The influence of the cantilever is negligible in the proximity of the measurement tip.

A part of the solely cantilever above the bus structure shows figure 2. The comparison between the force distributions shows clearly that interaction forces on a cantilever without a measurement tip are much greater than on a cantilever with a measurement tip. This means, that in case of conducting lines below the cantilever, additional forces make a contribution to the cantilever bending. In this case the forces have to be considered for the correct interpretation of the experimental results.

The real time bandwidth of a typical cantilever is limited to ~10 kHz because of its mechanical low pass oscillation behaviour. Therefore, the measurement of voltages with higher frequencies necessitates a sampling technique. The quadratic formula (1) makes sampling techniques possible. Important conditions for a suitable logic state mapping are a positive and negative voltage response and a wide voltage range with linear sensitivity. This is possible with special sample pulses in phase to the DUT voltage carrying the resonance frequency of the cantilever.

The block diagram of the used EFM-test system for logic state mapping of digital signals is shown in figure 3. The EFM-probe is scanned contactlessly in a constant distance to the conducting lines of the DUT. The DUT voltage $u_m(t,x,y)$ is applied by a data generator. Synchronously to the period $T$ of the DUT voltage a trigger signal is generated. This trigger signal is delayed by a scan generator signal and is shaped into a short sample pulse. The fast chop up circuit generates the required EFM-probe voltage $u_p(t)$ consisting of the sample pulses amplitude modulated with the cantilever resonance frequency

$$u_p(t)=P(t)\cdot\hat{u}_p\cdot\cos(\omega_{res}t) \tag{2}$$

where $P(t)$ is the shape of the pulse and $\hat{u}_p$ is the amplitude of the EFM-probe voltage. The trigger delay $\Delta t$ is not mentioned explicitly. The temporal mean value of $u_p(t)$ is zero. This was considered in the simulations.

The cantilever bending is analyzed via a lock-in amplifier. Its output is recorded by a computer control. The output is proportional to the resulting force at the cantilever resonance frequency $\omega_{res}$

$$F|_{\omega_{res}} \sim \left(\frac{1}{T}\int_0^T P(t)\cdot u_m(t)\ dt\right)\cdot\hat{u}_p\cdot\cos(\omega_{res}t)\,. \tag{3}$$

The interaction force amplitude is proportional to the term in brackets. This is a linear term depending on the unknown voltage. In that way a linear dependency between the force and $u_m(t)$ is achieved.

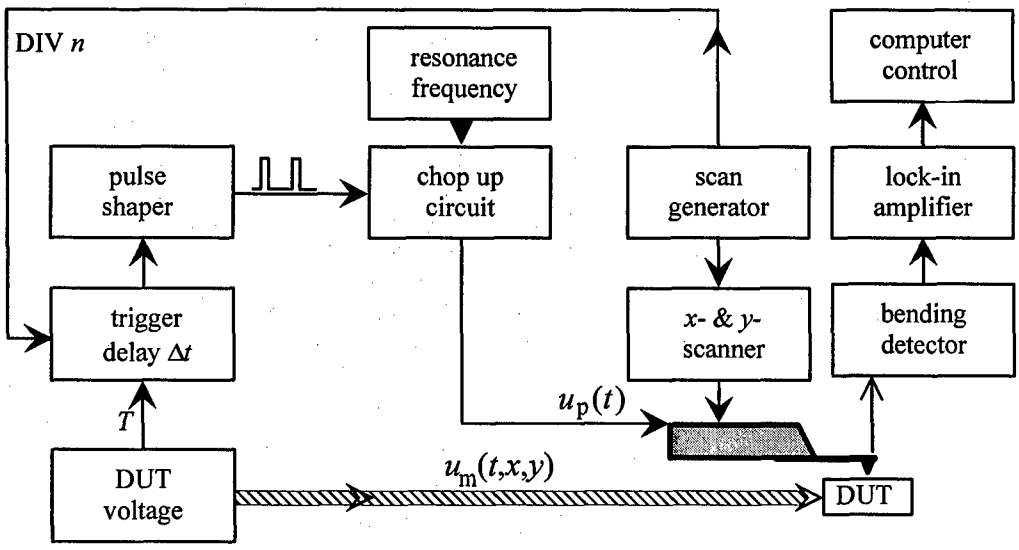



Fig. 3 Block diagram of the set-up for logic state mapping with an EFM

In this set-up the used pulse shaper is a monoflop. The generated pulses show a rectangular shape with a width of approximately 150 ns. A similar set-up for point measurements was reported recently [14]. The pulse shaper was a step-recovery diode which enables qualitative point measurements on special test pads with a pulse width of 350 ps. IC-internal measurements were not presented.

## 3. Measurement results

### *3.1 Test structure*

The experimental results of the contactless stroboscopic potential contrast and the logic state mapping were done on an electron beam direct written three line bus structure of an unpassivated ITG test chip [15]. The 2 μm wide conducting lines form the narrowest uninterrupted bus structure of this test chip. They show a spacing of 1 μm. They are surrounded by ground signals (GND). The applied signals on the conducting lines are 1 MHz (①,②) and 2 MHz (③) clock signals (see Fig. 4). The IC-operation states with a distance of 250 ns are marked.

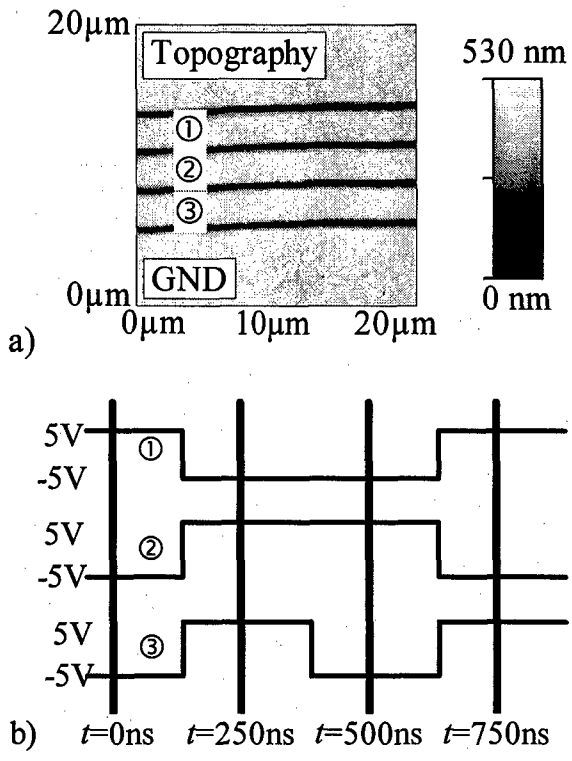



Fig. 4: Test structure: a) topographical image of the bus structure b) applied voltage with marked IC-operation states

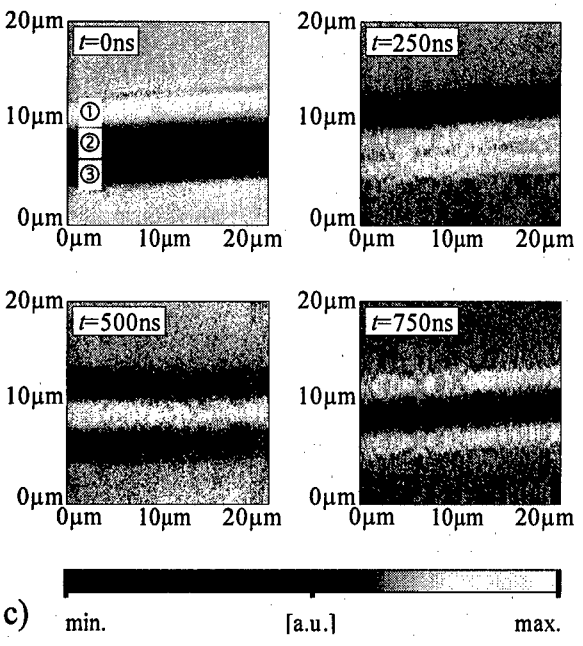



Fig. 5: Stroboscopic voltage contrast: quasistatic micrographs at different IC-operating states corresponding to the delay times according to Fig. 4b) (see insets), bright: +5 V, dark: -5 V

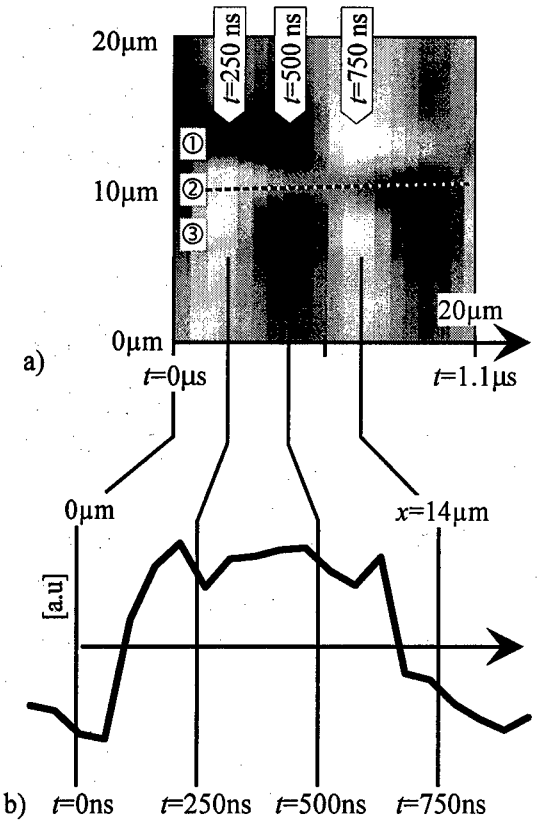



Fig. 6: Logic state mapping: a) on the bus structure b) deconvoluted line scan of the middle conducting line ②

### 3.2 Stroboscopic voltage contrast

The phase relation between the IC signal and the sampling pulse is fixed during scanning, so that a quasistatic micrograph (time slice) of a definite IC state can be recorded. By varying the trigger delay $\Delta t$ the voltage distribution at different IC-operation states are being displayed. It is possible to examine the logic states simultaneously with the lateral distribution of this states. This enables an orientation inside of the IC and a logic analysis of the IC-operation states.

The figure 5 shows a stroboscopic micrograph depicted with the realized set-up directly and without any subsequent image processing. The delay time $\Delta t$ was adjusted at the end of every full picture and was chosen identically with the required IC-operation state.

The good agreement with the expected voltage distribution of the IC-operation states demonstrates the functioning of the realized set-up. The conducting lines, the spaces, the logic levels and the ground voltage (GND) are clearly visible. Thus, the spatial resolution is much better then the width of the conducting lines. The logic levels at each IC-operation state can be determined easily.

### 3.3 Logic state mapping

For the logic state mapping the measurement tip scans in $y$-direction across a line perpendicular to the IC conducting lines and the trigger delay $\Delta t$ is incremented in the $x$-direction. In this way the $y$-direction represents only spatial information, whereas the $x$-direction represents the temporal and spatial information.

The figure 6 shows an example of a logic state mapping of the digital pattern of 1 & 2 MHz with the same conditions according to figure 4b). The trigger delay is incremented every 10 lines. The $x$-direction represents delay times up to 1.1 µs. The scan area is 20 µm X 20 µm. Despite of some noise

the logic levels and their timing are depicted quite precisely and a pattern recognition can be done easily, too. For example, the pattern at 250 ns is determined as ①: dark: -5 V, ②: bright: +5 V, ③: bright: +5 V. This is the expected IC-operating state for 250 ns (see figure 4).

## 4. Discussion

Basic simulations for the determination of the force distribution along the shape of a typical measurement tip showed that the proved spatial resolution is not a physical limit for EFM-testing. In fact, it is limited by the bus structure used here. A spatial resolution in the range of 100 nm can be expected from the simulation which is in agreement with previous experimental results.

The temporal resolution is limited by the realized set-up, especially by the monoflop used. It could be improved significantly by using a step-recovery diode as the pulse shaper.

## 5. Conclusion

A novel set-up for time resolved logic analysis was designed and adapted to an EFM-testing system. The functioning of the realized set-up was shown. The actual available performance is demonstrated by means of stroboscopic voltage contrast and logic state mapping at megahertz clock signals applied to a bus structure with 2 μm conducting lines. These testing techniques are now available for EFM-testing. Spatial and temporal resolution progress can be achieved by the realization of the discussed improvements.

## Acknowledgements

This work was financially supported by the MWF (ministry of science and research of North-Rhine-Westfalia) and by Siemens AG, Munich.

## References

[1] Menzel E., Microelectron. Eng. 16 (1992) 3-14 (and references)

[2] Schöttler M., Storzer H.D., and Kubalek E. Microelectron. Eng. 16 (1992) 497-504

[3] Martin Y., Abraham D.W., and Wickramasinghe H.K. Appl. Phys. Lett., 52 (1988), 1103-1105

[4] Bloom D.M. Microelectron. Eng. 24 (1994) 3-9

[5] Sprengepiel J., Böhm C., and Kubalek E. Microelectron. Eng. 31 (1996) 181-186

[6] Leyk A., Böhm C., and Kubalek E. IEEE MTT-S International Microwave Symposium Digest, Vol. 3 (1996) 1525-1528

[7] Sprengepiel J., Böhm C., and Kubalek E. in Proceedings of the 21th ESSCIRC '95, Lille, France, (1995) 11-17

[8] Bridges G.E., Said R.A., Mittal M., and Thomson D.J. J., Vac. Sci. Technol. B 13(3) (1995) 1375-1379

[9] Bangert J. and Kubalek E. Microelectron. Reliab. Vol 37 No. 10/11 (1997) 1579-1582

[10] Bridges G.E. and Thomson D.J. Ultramicroscopy 42-44 (1992) 321-328

[11] Wittpahl V., Behnke U., Wand B., and Mertin W., Microelectron. Reliab. (1998), published in this issue

[12] Linde H.J. van Math. Comp. 28 (1974) 369-391

[13] Müller U., Hoffschen S., BöhmC., Sprengepiel J., Kubalek E., and Beyer A. Microelectron. Eng. 31 (1996) 235-240

[14] Brigdes G.E., Noruttun D., Said R.A., Thomson D.J, Lam T., and Qi R. J. Vac. Sci. Technol. A 16(2) (1998), 830-833

[15] Richter H. Microelectron. Eng. 16 (1992) 225-232

PERGAMON

Microelectronics Reliability 38 (1998) 957–961

MICROELECTRONICS RELIABILITY

# Failure analysis of integrated devices by Scanning Thermal Microscopy (SThM)

G.B.M. Fiege[a], V. Feige[a], J.C.H Phang[b], M. Maywald[c], S. Görlich[c], L.J. Balk[a]

[a] *BUGH Wuppertal, Lehrstuhl für Elektronik, Wuppertal, Germany*
[b] *Centre for IC Failure Analysis and Reliability, National University of Singapore, Singapore*
[c] *Siemens AG, Semiconductor, Failure Analysis, Munich, Germany*

**Abstract**

High power densities dissipated in smaller and faster devices are leading to major thermal problems of semiconductor devices. The resulting local heat dissipation can induce deleterious effects like accelerated degradation or the destruction of the integrated circuits. Due to the shrinking feature sizes of modern devices and the small local extension of electrical failures the exact localization of these defects using established thermal failure analysis techniques like infrared thermometry becoming more and more difficult. Temperature measurements on passivated electronic devices with a sensitivity of 5 millikelvin by the use of a scanning thermal microscope (SThM) in contrast demonstrate the possibilities to use this system as a tool for failure analysis. Hot spot imaging with a spatial resolution of less than 150 nm, investigations on the backside of ULSI devices as well as a comparison with complementary established analysis techniques are presented.
© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

The large scale integration and the consequent reduction of feature sizes in microelectronics cause more and more problems for the exact localization of electrical failures. A wide-spread technique for failure analysis is infrared thermometry, based on the detection of the infrared light emitted from hot spots induced by current leakage or short circuits. The achievable resolution of this and other optically based techniques is diffraction limited to more than 0.3 μm.

Further developments of scanning probe microscope (SPM) based techniques in the last years show the potential of this powerful tool for the analysis of semiconducting materials and devices with a high spatial resolution. Techniques which enable the investigation of thermal properties in the nanometer regime were also presented.

The use of a microscopic thermocouple in combination with a scanning tunneling microscope (STM) allows recording of the local thermal properties of materials and devices - assuming a satisfactorily high electrical conductivity of the samples for the tunneling process [1]. This assumption is not necessary if the thermocouple is placed at the end of the tip of a scanning force microscope (SFM) because the distance control is force based and therefore independent of a current flow into the sample [2]. Temperature profiling with

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00086-9

high temperature and spatial resolution as well as the application of this technique to failure analysis are carried out by Majumdar et al. [3]. It is also possible to use a microscopic thermocouple for the investigation of the local thermal conductivity of a sample, but for these measurements it is necessary to additionally heat up the probe by an external heating source.

Our setup, based on a resistive thermal probe, on the other hand allows the measurement of the temperature distribution as well as the local thermal conductivity on a device under test without any additional heating source [4].

## 2. Experimental Setup

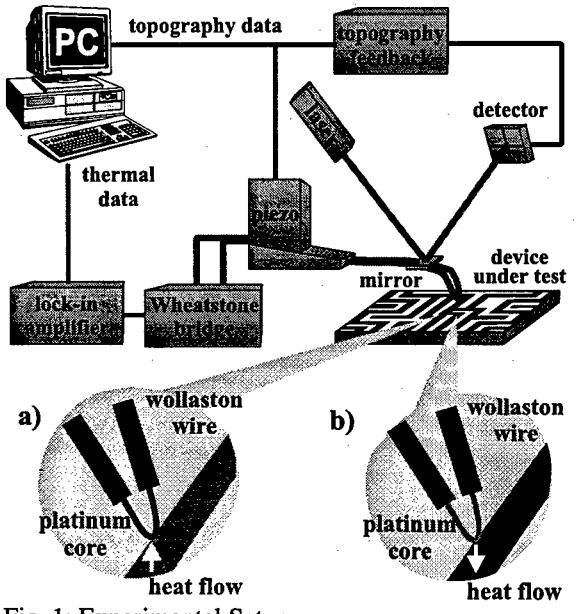



Fig. 1: Experimental Setup

Figure 1 shows the experimental setup of our Scanning Thermal Microscope (SThM).

For distance control a common scheme consisting of a laser and a four-cell photodetector is used. A small mirror fixed to the thermal probe on which the feedback laser beam is reflected allows the determination of the displacement of the cantilever during the scan. Information about the local temperature on the device under test can be obtained by measuring the probe resistance changes caused by variations in sample surface temperature (Fig. 1a). The operation in the constant force mode ensures a constant contact area during the image acquisition, therefore the measured resistance changes are proportional to the temperature changes of the sample surface.

To measure the local thermal conductivity of a device, a larger current is passed through the probe in order to induce resistive self heating. In this mode a heat flow from the probe into the sample takes place, this heat flow is influenced by the thermal conductivity of the sample(Fig. 1b). Local thermal conductivity is accessed by keeping the tip at a constant temperature while monitoring the heating power.

An optimization of the temperature detection scheme allowed to increase the resolution to 5 mK as calculated from noise analyses, which is one order of magnitude better than the thermal noise level of established techniques for failure analysis like the fluorescent microthermal imaging technique [5]. This calculated noise value could be confirmed by temperature measurements on a test structure consisting of heatable passivated metal lines.

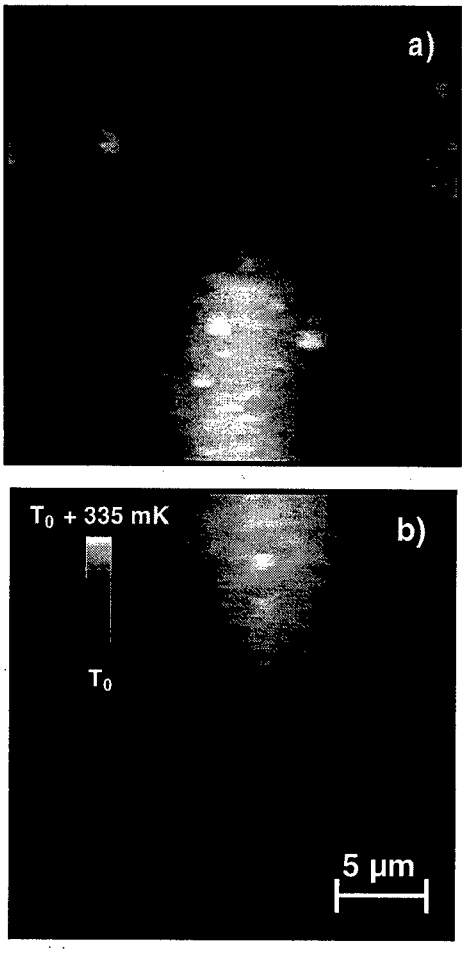



Fig. 2: Topography (a) and corresponding temperature micrograph (b) of a passivated tapered metal line

## 3. Results and Discussion

Figure 2 shows a temperature micrograph of a passivated tapered metal line on a silicon substrate heated up by a current flow (passivation layer thickness = 1.3 μm). The temperature increase in the upper part of the image as a result of the high current density ($J=2x10^{11}$ A/m$^2$ in comparison to a current density of $J=3x10^{10}$ A/m$^2$ at the 10 μm wide Au-line in the lower part of the image) is clearly visible.

To demonstrate the possibilities to use the system as a tool for failure analysis an n-channel MOSFET with gate-drain breakthrough was investigated. The device was biased at saturation ($U_{gs}$ = 3 V, $U_{ds}$ = 6.5 V), leading to a drain to gate current of 10 mA.

In Fig. 3 the location of the breakthrough can clearly be seen. The 1.2 μm thick passivation layer is cracked and forming a hill, visible on the left side of the topography image. The induced hot spot in the temperature micrograph is located below this hill. The noise-equivalent spatial resolution (NESR), defined as the ratio of the measured noise versus the largest temperature gradient [2], was determined to be better than 150 nm. This spatial resolution was calculated from the experimental data of one temperature measurement on a passivated structure and does not represent a fundamental limitation of the experiment setup. This definition of the achieved spatial resolution takes into consideration that structure and geometry of the sample as well as their electrical and thermal conditions strongly affect the achievable resolution of a thermal measurement. Due to the high power dissipation at the breakthrough region the observed high temperature gradients lead to a high spatial resolution. At devices with a nearly uniform temperature distribution and therefore small temperature gradients the observed spatial resolution will be reduced.

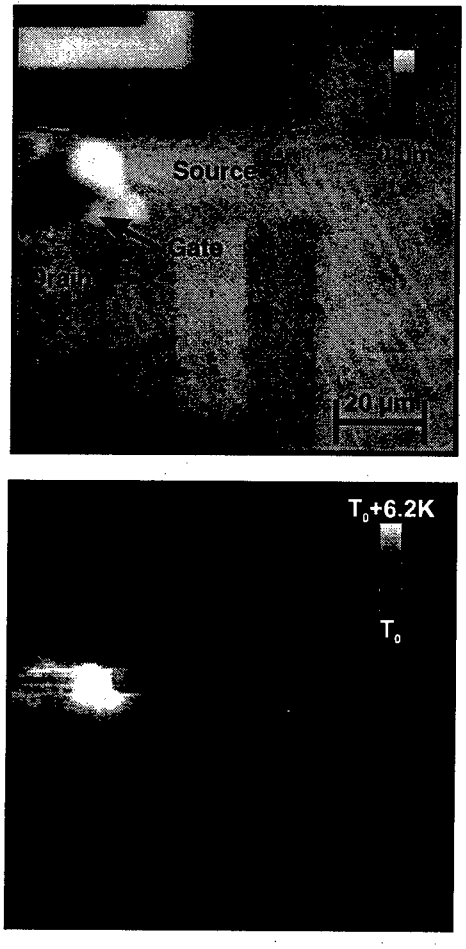



Fig. 3: Topography and temperature distribution of a gate-drain breakthrough of a MOSFET

The limiting factor of our system is the size of the area in which the heat flow from the sample into the probe takes place. From local thermal conductivity measurements this area and therefore the resolution of our system could be estimated to 30 nm [6], which is leading to the assumption that on appropriate samples a spatial resolution in the same range should be possible.

On the same type of n-channel MOSFET with 1.6 μm gate length comprehensive measurements with a spectroscopic photon emission microscope system (SPEMS) were carried out [7,8]. A typical I/V-curve of this device is illustrated in figure 4. The transistor was operated for both measurements at a gate voltage of 3V and a drain current of 4.3 mA.

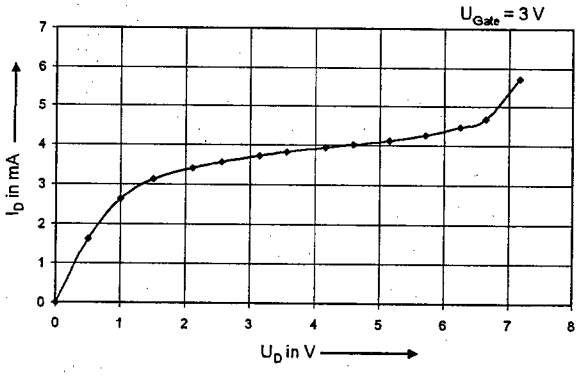



Fig. 4: I/V-curve of the n-channel MOSFET

Fig. 5 shows the topography, the corresponding temperature micrograph, and an overlay of an image

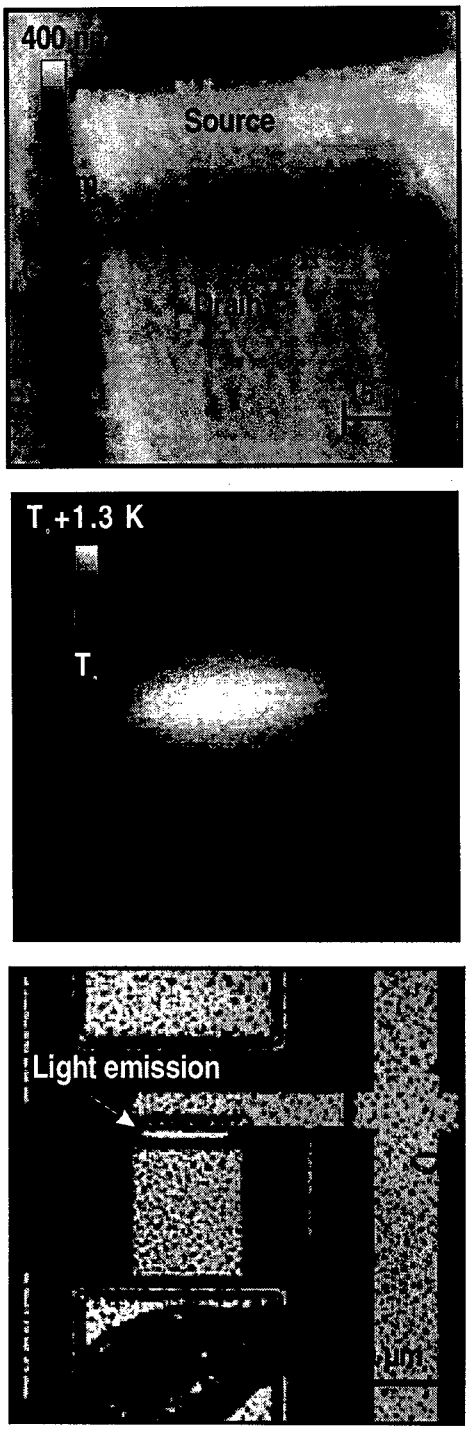



Fig. 5: Comparison of SThM and SPEM analyses of an n-channel MOSFET biased at saturation

obtained by illuminating the IC and a light emission image.

The injection of electrons from the channel into the highly resistive pinch-off region close to the drain leads to light emission in a long and narrow region at the edge close to the drain. The light emission is restricted to the area of the substrate between the opaque gate and drain metallizations. Unfortunately, by the SPEM technique it is not possible to obtain information about the substrate below the metallisation. However, in the SThM image the point of highest temperature is also located in this region close to the drain. In contrast to the SPEM technique the temperature distribution due to the good heat conduction of metal is also detectable in the metallized areas.

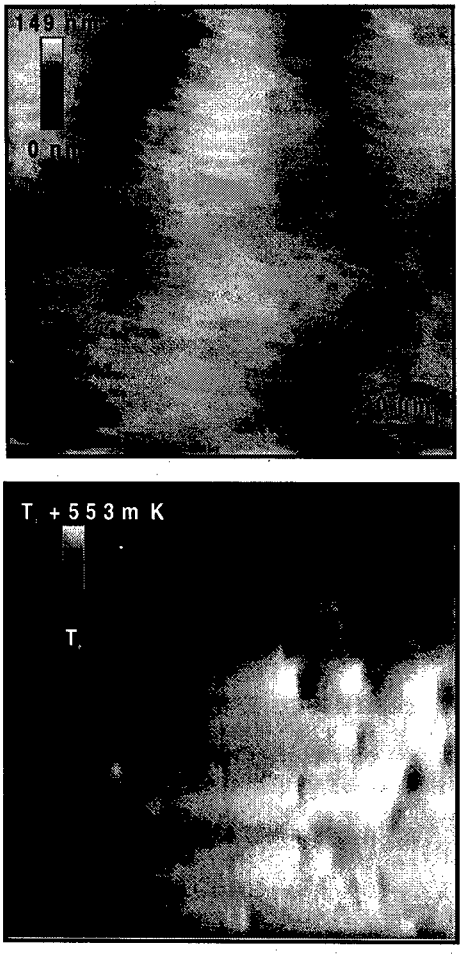



Fig. 6: Topography and Temperature distribution on the backside of a ULSI device

The large number of layers above the active structures of modern integrated circuits and the resulting big distance to the sample surface are leading to extensive problems for failure analysis

from the frontside. SThM as a probe technique, aquiring information of the physical properties at the sample surface, is also affected by this development. Therefore we carried out first experiments of temperature measurements on a thinned substrate of a ULSI device. In Fig. 6 a hot spot is clearly visible in the temperature micrograph recorded on the backside of the sample; the thickness of the substrate was reduced by polishing to 20 μm. The size of the hot spot exceeds the scan range of 100 x 100 μm as confirmed by complementary SPEM measurements.

## 4. Conclusions

The temperature resolution of our setup could be improved to 5 mK, as estimated by noise analyses and confirmed by temperature measurements on test structures. We have demonstrated that SThM is a suitable tool for failure analysis of integrated circuits. On passivated structures as well as on the backside of ULSI devices temperature measurements with a noise-equivalent spatial resolution of up to 150 nm have been presented.

## References

[1] Oesterschulze E, Stopka M, Ackermann L, Scholz W and Werner S. Thermal imaging of thin films by scanning thermal microscope. J. Vac. Sci. Technol. B14(2) (1996) 832-837.

[2] Majumdar A, Luo K, Shi Z and Varesi J. Scanning thermal microscopy at nanometer scales: A new frontier in experimental heat transfer. Experimental Heat Transfer 9 (1996) 83-103.

[3] Lai J, Chandrachood M, Majumdar A and Carrejo J P. Thermal detection of device failure by atomic force microscopy. IEEE Electron Device Letters 16 (1995) 312-315.

[4] Cramer R M, Bruchhaus L, Balk L J. Characterization of Thermal Device Properties with Nanometer Resolution. Microelectron. Reliab. Vol. 37 (1997) 1583-1586.

[5] Tangyunyong P, Liang A Y, Righter A W, Barton D L and Soden J M. Localizing Heat-Generation Defects Using Fluorescent Microthermal Imaging. Proceedings of the 22nd ISTFA (1996) 55-62.

[6] Balk L J, Maywald M and Pylkki R J, Nanoscopic detection of the thermal conductivity of compound semiconductor materials by enhanced scanning thermal microscopy. Inst. Phys. Conf. Ser. No. 146 (1995) 655-658.

[7] Tao J M, Liu Y Y, Chan D S H, Phang J C H and Cim W K. A New Spectroscopic Photon Emission Microscope for Semiconductor Device Analysis. Proceedings of the 5th IPFA (1995) 60-65.

[8] Tao J M, Chim W K, Chan D S H, Phang J C H, Liu Y Y. A high-sensitivity photon emission microscope system with continuous wavelength spectroscopic capability. Proceedings of the 1996 IEEE International Reliability Physics Symposium, (1996) 360-365.

PERGAMON

Microelectronics Reliability 38 (1998) 963–968

MICROELECTRONICS RELIABILITY

# Scanning near-field optical microscopy analyses of electronic devices

R.M. Cramer[a], W.R. Schade[a], R. Heiderhoff[a], L.J. Balk[a], R. Chin[b]

[a] *Lehrstuhl für Elektronik, Bergische Universität Wuppertal, 42097 Wuppertal, Germany*
[b] *Intel Corp., 5200 N.E. Elam Young Pkway, Hillsboro, OR 97124, U.S.A.*

**Abstract**

We discuss the usefulness of scanning near-field optical microscopy (SNOM) techniques for the optical failure analysis of electronic devices with a lateral resolution well below the diffraction limit. By imaging dielectric contrasts, defects within metallization layers can be located while optical near-field induced current/conductivity (ONIC/ONICond) analyses allow the characterization of device internal electronical properties with less than 50 nm resolution. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

In the recent past optical techniques such as laser scanning microscopy (LSM) and optical beam induced current (OBIC) investigations have proven themselves as indispensable tools for the failure analysis of ultra-large scale integrated circuits and microwave devices, although their lateral resolving power is limited by the diffraction limit or Abbé-barrier to a few hundred nanometers [1]. For the oncoming process generations the requirements of the continuing tendency for miniaturization may be met by switching to UV radiation as the source of illumination, but a fundamental limit will be reached when today's low-k dielectrics become opaque due to the excitation of electron-hole-pairs within the inter-layer dielectrics (ILD). A solution to this dilemma may be presented by the introduction of scanning near-field optical microscopy (SNOM), where a nanometrical aperture is scanned in close proximity, i.e. a few nanometers, to the sample surface. Since the illumination of the sample takes place within the optical near-field, the achievable spatial resolution is only limited by the aperture's diameter and its distance to the sample rather than by the optical wavelength utilized for the analysis [2,3]. Although the theory for near-field optics has already been proposed by Synge in 1928 [4], its realization has not been practical until the introduction of the scanning probe microscope. Since almost any contrast mechanism used in conventional optical microscopy can be exploited at a higher resolution in SNOM [5], a tremendous amount of work has been invested into this area over the last years. As a result the amount of publications on SNOM currently outnumbers the amount of publications on scanning force microscopy (SFM) by far.

## 2. Experimental setup

As it turned out that SNOM probes based on optical fibers at the moment show improved

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00082-1

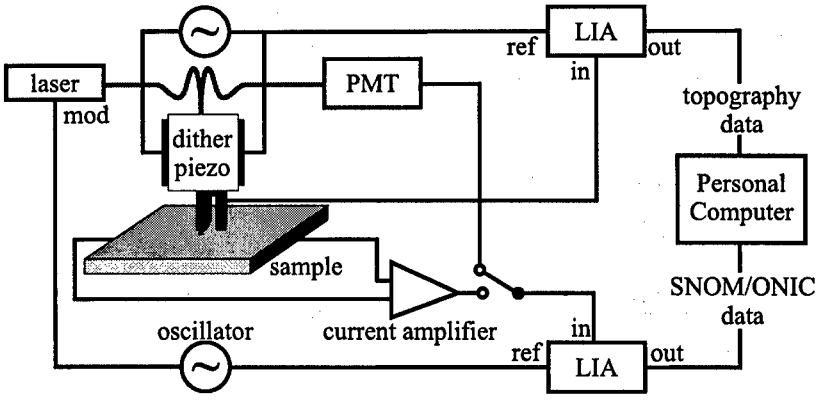



Fig. 1: Schematical illustration of the experimental setup

transmission characteristics in comparison to micromachined aperture probes, we have adopted this configuration into our modified TopoMetrix Aurora SNOM in conjunction with a distance regulation scheme where the Al-coated tip is glued to one of the two prongs of a piezoelectric tuning fork [6]. The tip/tuning fork system is dithered laterally over the sample surface at a typical amplitude of 3..5 nm in such a way that the tuning fork's electrical output signal represents the oscillation amplitude. As the tip approaches the surface, the oscillation is damped by shear-forces and the demodulated tuning fork signal can be used as a highly sensitive means for the distance regulation. By analyzing the phase signal, we have also been able to observe a material contrast, as recently discussed by Williamson et. al. [7]. With this technique we can achieve height resolutions in the sub-nanometer range and tip lifetimes of several ten measurements. Fig. 1 shows the experimental setup used in this work.

The devised instrument is intended to be used in failure analysis labs by operators on a technician level. Therefore, the criteria for the realization included ease-of-use, handling and preparation times of both sample and instrument. To improve the ease-of use, an automated procedure for setting the tuning fork feedback parameters has been developed by using a programmable lock-in-amplifier (LIA). This procedure automatically adjusts both excitation frequency and lock-in phase to the optimum values for the subsequent scanning process. As it has been fully integrated into the scanning software (TopoMetrix SPMTools and SPMLab), the instrument can be operated with a minimum amount of training, as opposed to many other SNOM setups which demand major knowledge of the physical processes and interactions in shear-force microscopy.

To simplify the sample navigation and to reveal information on the remaining passivation layer thickness of the sample after deprocessing, a high-resolution color CCD-camera has been implemented into the setup, the resulting magnification allows us to navigate the SNOM probe with an accuracy of a few cells.

The original instrument permits reflection as well as transmission mode SNOM. Since at this time the utilized radiation for both dielectric contrast imaging and OBIC analyses is in the visible range, the transmission optics have been replaced by an improved sample fixture for semiconductor samples which can easily hold standard scanning electron microscopy sample holders and a vast variety of packaged devices. Finally, the Argon-ion laser has been replaced by a fiber coupled diode laser to avoid frequent adjustments and to allow for a light modulation at up to 10 MHz.

## 3. Results

### *3.1. Dielectric contrast imaging*

It was recently demonstrated that SNOM can be used to generate optical images of integrated devices with a lateral resolution of well below 50 nm [8]. For example, Fig. 2 shows a comparison of a LSM image (Zeiss LSM, utilized wavelength $\lambda = 488$ nm) and a SNOM micrograph of a focused ion beam (FIB) deprocessed SRAM cell array at the first metal layer. The SNOM micrograph has been recorded in reflection by an external objective at an angle of 45° with respect to the optical axis. Since the SNOM

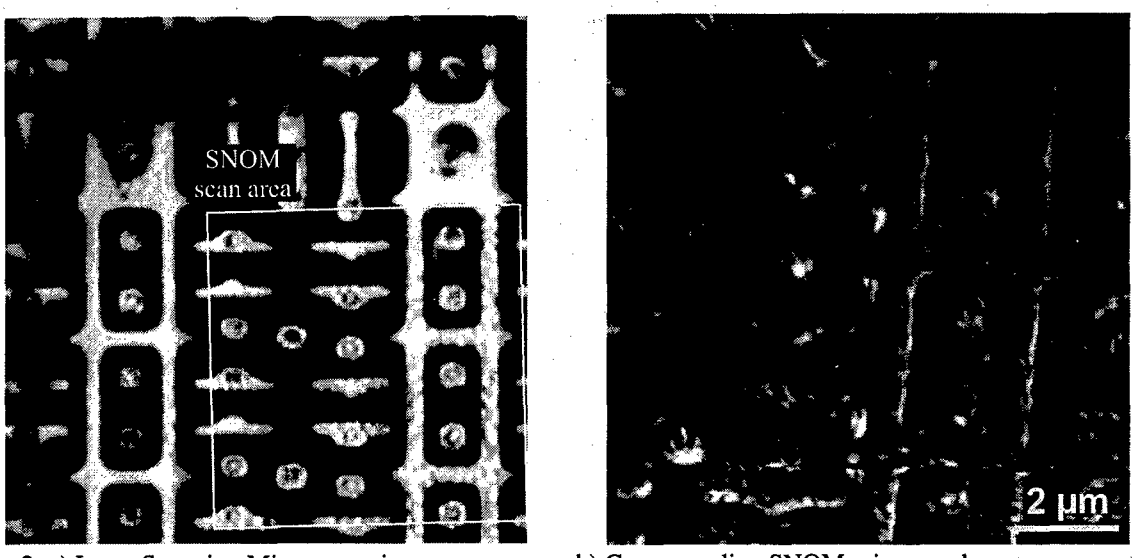



Fig. 2: a) Laser Scanning Microscopy image b) Corresponding SNOM micrograph

scan range can further be decreased by a factor of ten, the gain in optical magnification becomes clearly visible.

Despite the fact that the achievable lateral resolution approaches the molecular scale, the actual interpretation of images is obstructed by detrimental effects. From Fig. 2b it can be seen that the contrast between the metal structures and the ILD is very poor and that additional features within the ILD can be seen which do not appear in the LSM image and would complicate a localization of defects. In our opinion shadowing of the reflected radiation by the probe itself caused by the commonly used 45° (off-axis) detection scheme is responsible for this effect [8]. To overcome this, we both illuminate the sample and subsequently detect the reflected light through the tip (on-axis). Since it has been previously reported that by using uncoated tips the resulting images are predominated by far-field rather than by near-field effects [9], we use coated fibers for our experiments. However, in this case the attenuation at the aperture is quite dramatic, fiber-internal reflection is taking place and the reflected signal has to pass the aperture twice. To be able to generate images despite these obstructions, highly sensitive techniques for the detection of the reflected signal must be applied using photomultiplier tubes and a lock-in amplification. In addition we have modified some tips after a method introduced by Veerman et. al. [10], where the very end of the fiber is being cut off using FIB. The resulting aperture is free of aluminum contaminations and perfectly circular so that the attenuation is one order of magnitude lower than without applying the FIB process. Fig. 3 shows by a comparison of topography, off-axis and on-axis SNOM micrographs of a freestanding aluminum pad how the interpretability of SNOM images can be enhanced by this detection scheme. The sample was prepared by a selective deprocessing of the ILD to generate a contrast situation which can theoretically be most easily evaluated, i.e. a large amount of reflection above the aluminum versus virtually no reflection when the tip is located above the air gap.

It becomes obvious that in the on-axis detection image the aluminum pad is much more clearly visible and has a higher contrast than in the off-axis micrograph, although the dynamic range of Fig. 3b) exceeds the dynamic range of Fig. 3c) by three orders of magnitude (the gray-scale in the on-axis micrograph has been adapted to account for the lower dynamic range). We believe that this behavior in the off-axis detection scheme is again caused by shading effects. While the tip is located above the aluminum pad, virtually all reflected light is blocked by the tip itself and only a small fraction can reach the detector. Above the air gap, however, no blocking can occur and the light both emitted and diffracted at the aperture and reflected in the far-field from the substrate can be picked up by the photomultiplier. In contrast, in the on-axis detection micrograph a large signal background from reflections inside the fiber is present but remains constant throughout the measurement, the dynamic range is thus extremely low. Any change in the signal

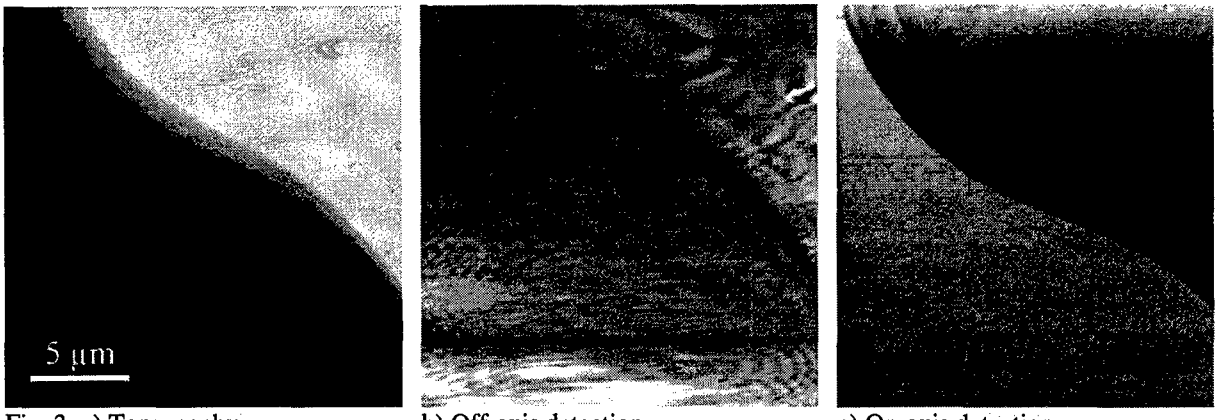



Fig. 3: a) Topography b) Off-axis detection c) On-axis detection

intensity on the other hand must be originating from a change in the complex dielectric constant of the material located directly underneath the aperture. However, the origin of the inverted contrast yet remains to be discussed. One would expect that the reflected signal is higher when scanning above the aluminum because of its high reflectivity, but surprisingly the opposite can be observed. As the aperture acts like a dipole source, we believe that the fields of this dipole and the mirror dipole which is induced inside the aluminum extinguish each other when detected in the far-field. An indication of this theory is that when the subwavelength aperture is exchanged by an aperture with a diameter of about 800 nm, which can be treated by conventional optics theories, the contrast situation reverses again. However, this effect is not yet fully understood and currently subject to extensive investigations, especially as the so-called z-motion artifact [11] could also account for the contrast inversion.

Also, since the dynamic range is extremely low in comparison to the background signal, only a limited amount of SNOM probes has shown the capability to perform an on-axis detection on this particular sample. When using samples with lower variances in the dielectric constant (e.g. passivated samples), the resulting micrographs cannot reveal optical information, thus the transmittivity of SNOM probes has to be increased in the future. A variety of promising approaches is conducted at the moment including micromachining so that new probes can be expected within the next years [12,13,14,15,16,17].

### *3.2. Optical near-field induced current/conductivity analyses*

In the same setup, optical near-field induced current/conductivity (ONIC) analyses can be carried out by locally generating electron-hole pairs in the near-field and externally measuring the induced current [18,19]. Optical near-field induced current investigations take advantage of device internal potential gradients leading to a separation of the induced electron-hole-pairs while an additional voltage is applied along the device under test for optical near-field conductivity (ONICond) analyses. The main advantage of this technique over conventional OBIC techniques, where the sample is locally excited by a laser beam, clearly is the gain in lateral resolution. To overcome the optical diffraction limit the EBIC (electron beam induced current) technique could also be applied, although the energy dissipation volume and the diffusion length will limit the achievable resolution as the electron beam penetrates deeply into the sample [20]. Other tradeoffs are the inevitable affection of the device internal electrical properties by the injected electrons and the need for placing the specimen into the vacuum chamber. STM-EBIC in contrast can improve the achievable spatial resolution, but again the device internal field distributions will be influenced by the injected electrons and the tip voltage [20], besides the fact that this method is impracticable for passivated devices [21].

The layer structure of the sample which has been analyzed using the ONICond technique is depicted in Fig. 4a), the motivation for the analyses has been a performance degradation after the gate recess process [22]. Fig. 4b) shows the topography

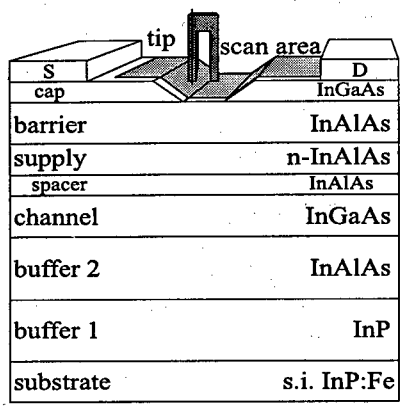



Fig. 4: a) HFET layer structure




b) Topography



c) ONIC micrograph

of the device, the height sensitivity of the instrument becomes obvious as the full grayscale represents a height variation of about 80 nm. In the ONICond micrograph, however, inhomogeneities within the electronic properties of the DUT become clearly visible which are likely to be introduced during the gate recess process. Besides the difference in the measured ONICond currents between the top and the bottom of the micrograph, several dark spots, which could be accounted for by recombination centers, are visible in the top right corner of the image. The achieved spatial resolution, which has been determined by a line analysis over the gate recess channel and the small hillocks, is below 50 nm. For further analyses of the defect characteristics spectroscopic and time-resolved analyses will gain information of the physical nature of the inhomogeneities.

## 4. Conclusions

Scanning near-field optical microscopy can be applied to generate optical images of passivated semiconductor samples at a resolution of less than 50 nanometers while the interpretability of the achieved results can be improved by using on-axis detection methods. The realized system can be operated without major practice and is optimized for sample handling and setup times. Although the transmittivity of the currently available tips does not allow for an imaging of arbitrary samples, recent developments of high-throughput SNOM probes indicate that this limitation can be overcome within the next years.

With the ONIC(ond) technique inhomogeneities within the electronical properties of HFET devices after a gate recess process have also been analyzed with 50 nm lateral resolution in the same system, showing the versatility of the realized microscope. By modulating the laser source and using lock-in techniques for a detection of the ONICond currents, time-resolved measurements are practicable as well as spectroscopic measurements, which can be performed by tuning the laser source.

## Acknowledgments

The authors would like to thank W. Daumann and F.-J. Tegude, Department of Solid State Electronics at the University of Duisburg, for supplying the HFET device.

## References

[1] Abbe E. Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. Archiv f. Mikroskop. Anat. **9** (1873) 413.

[2] Ash E A and Nicholls G. Super-resolution Aperture Scanning Microscopy. Nature **237** (1972) 510-512.

[3] Pohl D W, Denk W and Lanz M. Optical Stethoscopy: Image recording with resolution λ/20. Appl. Phys. Lett. **60** (24) (1984) 2957-2959.

[4] Synge E H. A Suggested Model for extending Microscopic Resolution into the Ultra-Microscopic Region. Philos. Mag. **6** (1928) 356.

[5] Betzig E and Trautman J K. Near-Field Optics: Microscopy, Spectroscopy, and Surface Modification Beyond the Diffraction Limit. Science **257** (1992) 189-195.

[6] Karrai K and Grober R D. Piezoelectric tip-sample distance control for near field optical microscopes. Appl. Phys. Lett. **66** (14) (1994) 1842-1844.

[7] Williamson R L, Brereton L J, Antognozzi M and Miles M J. Are artefacts in scanning near-field optical microscopy related to the misuse of shear force? Ultramicroscopy **71** (1998) 165-175.

[8] Cramer R M, Balk L J, Chin R, Boylan R, Kammer S B, Reineke F J and Utlaut M. The use of near-field scanning optical microscopy for failure analysis of ULSI circuits. ISTFA '96.Proceedings of the 22nd International Symposium for Testing and Failure Analysis (1996) 19-24

[9] Sandoghdar V, Wegscheider S, Krausch G and Mlynek J. Reflection scanning near-field optical microscopy with uncoated fiber tips: how good is the resolution really? J. Appl. Phys. **81** (6) (1997) 2499-2503.

[10] Veerman J A, Otter A M, Kuipers L and van Hulst N F. High definition aperture probes for Near-field Optical Microscopy fabricated by Focused Ion Beam Milling. Appl. Phys. Lett. **72** (24) (1998) 3115-3117.

[11] Hecht B, Bielefeldt H, Inouye Y, Pohl D W and Novotny L. Facts and artifacts in near-field optical microscopy. J. Appl. Phys. **81** (6) (1997) 2492-2498.

[12] Abraham M, Ehrfeld W, Lacher M, Mayr K, Noell W, Guthner P and Barenz J. Micromachined aperture probe tip for multifunctional scanning probe microscopy. Proc. SPIE - Int. Soc. Opt. Eng. **3009** (1997) 34-42.

[13] Bukofsky S J and Grober R D. Video rate near-field scanning optical microscopy. Appl. Phys. Lett. **71** (19) (1997) 2749-2751.

[14] Essaidi N, Chen Y, Kottler V, Cambril E, Mayeux C, Ronarch N and Vieu C. Fabrication and characterisation of optical-fiber nanoprobes for scanning near-field optical microscopy. Appl. Opt. **37** (4) (1998) 609-615.

[15] Islam M N, Zhao X K, Said A A, Mickel S S and Vail C F. High-efficiency and high-resolution fiber-optic probes for near field imaging and spectroscopy. Appl. Phys. Lett. **71** (20) (1997) 2886-2888.

[16] Mononobe S and Ohtsu M. Development of a fiber used for fabricating application oriented near-field optical probes. IEEE Photonics Technol. Lett. **10** (1) (1998) 99-101.

[17] Munster S, Werner S, Mihalcea C, Scholz W and Oesterschulze E. Novel micromachined cantilever sensors for scanning near-field optical microscopy. J. Microsc. **186** (1997) 17-22.

[18] Buratto S K, Hsu J W, Betzig E, Trautman J K, Bylsma R B, Bahr C C and Cardillo M J. Near-field photoconductivity: application to carrier transport in InGaAsP quantum well lasers. Appl. Phys. Lett. **65** (21) (1994) 2654-2656.

[19] Cramer R M, Heiderhoff R, Selbeck J and Balk L J. Advanced Scanning Near-Field Optical Microscopy of Semiconducting Materials and Devices. Inst. of Phys. Conf. Ser. **157** (1997) 685-688.

[20] Koschinski P, Dworak V, West P and Balk L J. Development of an STM-based electron beam induced current (EBIC) microscope. J. Scanning Micr. **10** (1) (1996) 33-38.

[21] Koschinski P, Fiege G B M and Balk L J. Nano-Field-Effect Microscopy of Electrical Inhomogeneities on InGaAs-Surfaces. Inst. of Phys. Conf. Ser. **146** (1995) 659-662.

[22] Daumann W, Scheffer F, Prost W and Tegude F-J. High power InAlAs/InGaAs/InP-HFET grown by MOVPE. (1997) 1-4

PERGAMON

Microelectronics Reliability 38 (1998) 969–974

MICROELECTRONICS RELIABILITY

# A new test method for contactless quantitative current measurement via scanning magneto-resistive probe microscopy

S. Bae[a], A. Schlensog[a], W. Mertin[a], E. Kubalek[a] and M. Maywald[b]

[a] *Werkstoffe der Elektrotechnik, Gerhard-Mercator-Universität Duisburg, Bismarckstr. 81, 47048 Duisburg, Germany*
[b] *Siemens AG, HL FA3, Balanstr. 73, 81541 München, Germany*

**Abstract**

The aim of this paper is to present a new test method for contactless quantitative current measurements via scanning magneto-resistive probe microscopy. Its basic principle is based on the detection of the magnetic field over a current-carrying conducting line by a magneto-resistive probe, which is fastened to a scanning force microscope (SFM). The function and efficiency of this method is demonstrated by measuring the current flow through a conducting line with hundred kilohertz bandwidth. Regarding to the function and failure analysis of integrated circuits the size of the magneto-resistive probe is reduced, and the obtained experimental results are discussed. © 1998 Elsevier Science Ltd. All rights reserved.

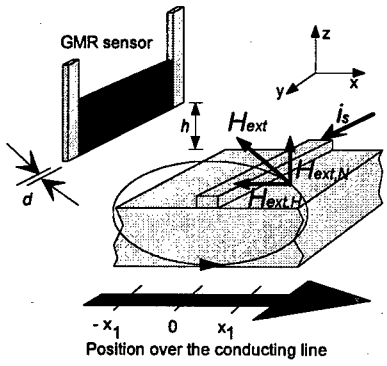
## 1. Introduction

The recent developments of an integrated circuit (IC) are characterized by increasing package densities and operating frequencies. Corresponding to this trend efficient internal test systems are required for function and failure analysis. Although different contactless test systems [1,2] for measuring internal voltages are present, additionally a test system to measure internal current quantitatively is still necessary for a more detailed analysis, e.g. investigating leakage currents or power consumptions in IC.

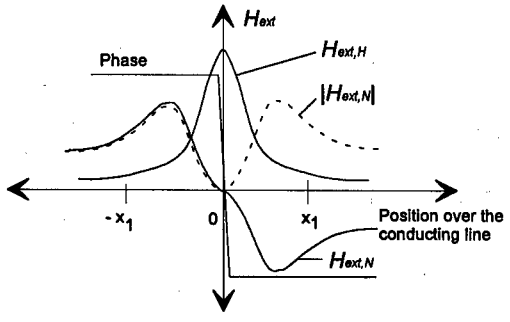
A test system for detection of internal currents can be realized by using different physical principles, e.g. the characterization of temperature [3][4] or magnetic fields [5,6,7] caused by a current flow through a conducting line. For quantitative current measurements it is advantageous to use test systems based on the characterization of magnetic fields, because test systems based on another physical principle require more extended material informations about the investigated test point. Therefore, only these test systems are discussed in this work.

With a magnetic force microscope the detection of internal currents with a resolution of 1 μA was already reported [5]. But the operating frequency range of the test system is limited by a resonance frequency of the used cantilever. Additionally, the quantitative current measurement is difficult, because the measured signal depends not only on the magnitude but also on the frequency of the detected current. With a test system based on a scanning electron microscope the detection of currents was performed [6]. However, its experimental results showed a low sensitivity of a detectable current of 100 mA. A newly developed scanning squid microscope (SSM) permits a detection of a smallest magnetic field at room temperature [7]. But it

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00071-7




(a)




(b)

Fig. 1 (a) Simulated magnetic field $H_{ext}$ outside the cross-section of a current-carrying conducting line, and (b) its horizontal component $H_{ext,H}$ and normal component $H_{ext,N}$ for a constant height $h$ over the conducting line and the phase of measured resistivity changes $\Delta R$.

is very complex and expensive. Until now no quantitative current measurements with a SSM are shown.

In this work a test system based on a magneto-resistive probe is realized and introduced. The used magneto-resistive probe offers the possibilities of the detection of currents down to μA range with megahertz band width [8] and, the realization of quantitative current measurements.

## 2. Measurement principle

The measurement principle of the scanning magneto-resistive probe microscopy is based on the detection of the magnetic field $H_{ext}$ caused by current flow $i_s$ through a conducting line. The magnetic field $H_{ext}$ outside the cross-section of the conducting line consists of a normal component $H_{ext,N}$ in z-direction and a horizontal component $H_{ext,H}$ in x-direction (see fig. 1.(a)). Both components of the magnetic field are shown in figure 1.(b) for a constant height $h$ over the conducting line.

To detect the magnetic field a giant-magneto-resistive (GMR) sensor is used as a magneto - resistive probe in this work (see fig. 1.(a)). Due to the very small thickness $d$ only the normal component $H_{ext,N}$ in z-direction can rotate the internal magnetization, which lies in y-direction and causes a resistivity change $\Delta R$ of the GMR sensor [9]. The resistivity change $\Delta R$ is proportional to the detected normal component $H_{ext,N}$.

The GMR sensor is positioned as shown in Figure 1.(a) and scans over the conducting line, which is fed with a direct current. The distribution of $H_{ext,N}$ can be detected by measuring the electrical resistivity change $\Delta R$ of the GMR sensor during the line scan (see fig. 1.(b)).

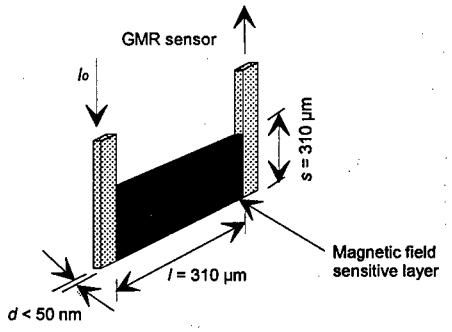
For an alternating current flow through a conducting line the measured electrical resistivity change $\Delta R$ corresponds to the absolute normal component $|H_{ext,N}|$. During the line scan the measured electrical resistivity change $\Delta R$ is zero in the middle of the current-carrying conducting line, and a 180° phase shift of $\Delta R$ occurs at this point (see fig. 1.(b)).

The measured electrical resistivity change $\Delta R$ is dependent on the scan height $h$ of a GMR sensor over a conducting line. Therefore, an unknown current flow through a conducting line with an unknown cross-section can be only quantitatively determined from the measured electrical resistivity changes $\Delta R$, if the topography of the cross-section of a conducting line can be measured and reproducible adjustments of the height $h$ of the GMR sensor over the conducting line are possible.

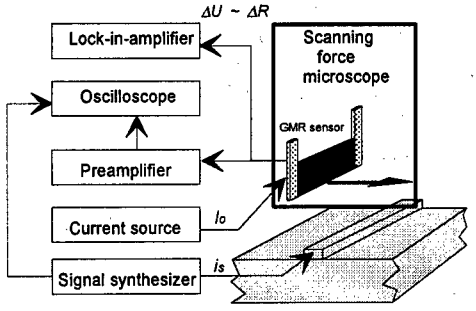
## 3. Measurement set-up

Figure 2 shows the measurement set-up for qualitative current measurements. A GMR sensor used for the detection of currents is shown in figure 2.(a). The GMR sensor is composed of multiple thin films with a thickness d < 50 nm. The magnetic field sensitive layer with an area of 310 μm x 310 μm corresponds to length $l$ and height $s$ of the GMR sensor.

The measurement set-up is described by the block diagram (see fig. 2.(b)). For a line scan over a current-carrying conducting line the magneto-resistive probe consisting of the GMR sensor was fastened to a SFM. The used SFM offered a




(a)




(b)

Fig. 2 Measurement set-up for qualitative current measurement : (a) principle of a used GMR sensor for the detection of currents, and (b) measurement set-up.

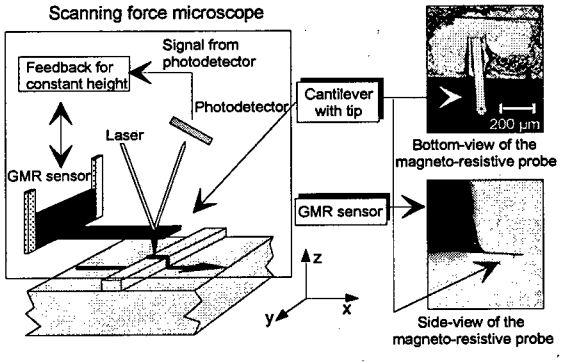



Fig. 3 Measurement set-up for quantitative current measurements : magneto-resistive probe consisting of a GMR sensor and a cantilever with a tip.

maximum line scan range of 150 µm. The current $i_s$ through the conducting line was generated by a signal synthesizer. The magneto-resistive probe was fed by a constant current $I_0$ and scanned over the conducting line. The measured electrical resistivity change $\Delta R$ causes a proportional voltage change $|\Delta U|$, which was analysed by a lock-in-amplifier. Simultaneously, the relative phase of $\Delta R$ was measured.

For quantitative current measurements we used a new magneto-resistive probe explained in figure 3. It consists of a commercially available cantilever with a tip, which is attached to the bottom of a GMR sensor. This offers now the possibility to investigate the topography.

## 4 Experiments and results

### *4.1 Qualitative current measurement*

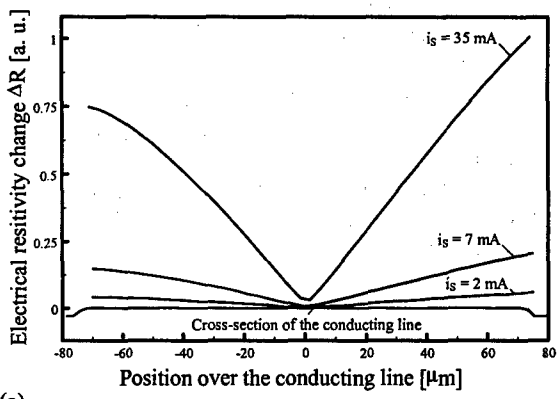
To demonstrate the function ability of the test system the detection of currents was performed on a 150 µm wide and 0.3 µm high conducting line, which was fed with currents $i_s$ between 35 mA and 2 mA with a frequency $f$ of 1 kHz (all magnitudes of the currents mentioned here are effective values). The line scans of the measured resistivity change $\Delta R$ are shown in figure 4.(a) with a measured topography of the cross-section of the conducting line. The magnitude of $\Delta R$ decreases proportional to the magnitude of the detected current.

To demonstrate the current sensitivity of the GMR sensor a current flow of 20 µA with a frequency $f$ of 1 kHz was detected on a 4 µm wide and 0.5 µm high conducting line (see fig. 4.(b)). Only the middle conducting line was fed with the current. In all these measurements it is obvious, that the line scans of $\Delta R$ show an asymmetry around the middle of the conducting line.

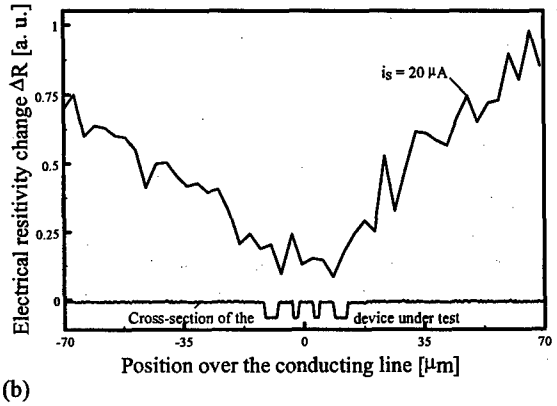
To investigate the frequency dependence of the GMR sensor a 150 µm wide and 0.3 µm high conducting line was fed with currents $i_s$ of 2 mA with a frequency $f$ between 1 kHz and 50 kHz (see fig. 4.(c)). The offset of the line scans of $\Delta R$ increases with increasing frequency $f$ of the detected current. The obtained line scan of $\Delta R$ for the frequency $f$ of 50 kHz is nearly flat. It is expected, that a parasitic signal from the current-carrying conducting line induces this offset and additionally it causes the above mentioned asymmetry of the line scans of $\Delta R$.

Therefore, to be able to measure higher frequencies the GMR sensor was electrically shielded. Additionally, to reduce the offset and the asymmetry a recalculation procedure was used.

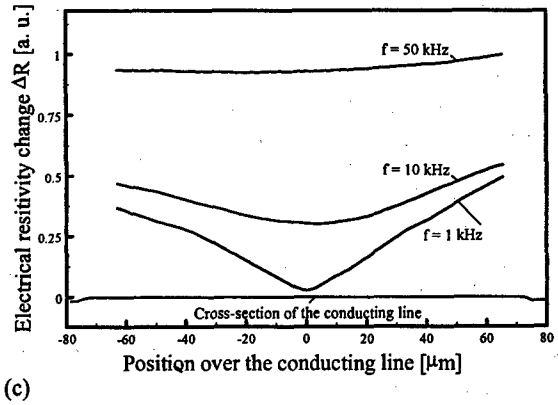
It is possible to select the offset and the asymmetry from the measured signal $\Delta R$ if its




(a)




(b)




(c)

Fig. 4 (a) Line scans of $\Delta R$ over a 150 μm wide and 0.3 μm high conducting line, which was fed with currents $i_s$ between 2 mA and 35 mA with frequency $f$ of 1 kHz, (b) line scan of $\Delta R$ over a 4 μm wide and 0.5 μm high conducting line, which was fed with $i_s$ = 20 μA with $f$ = 1 kHz, and (c) frequency dependence of line scans of $\Delta R$ over a 150 μm wide and 0.3 high conducting line, which was fed with $i_s$ = 2 mA with $f$ between 1 kHz and 50 kHz.

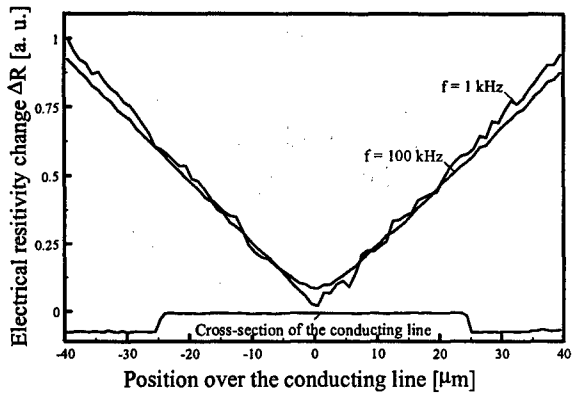
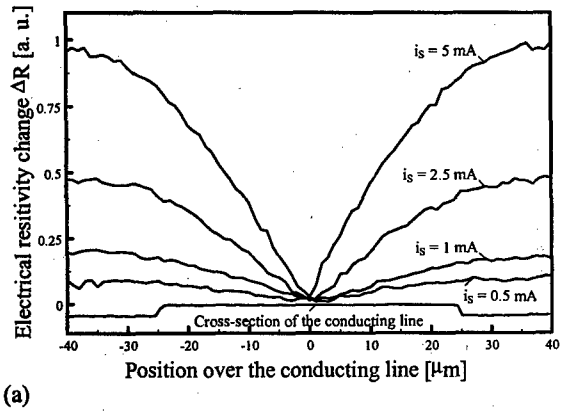
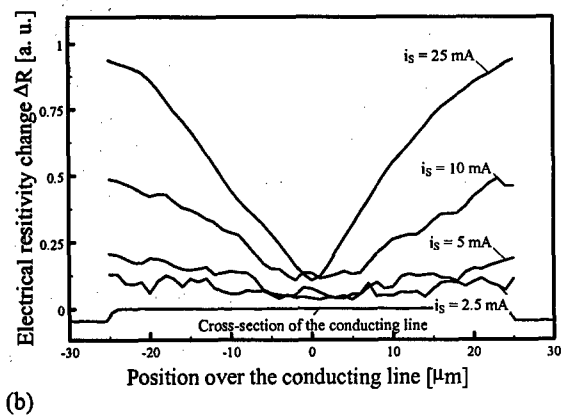



Fig. 5 Comparison between the recalculated line scan of $\Delta R$ from the detected current $i_s$ of 10 mA with $f$ = 100 kHz and the measured line scan for $f$ = 1 kHz




(a)




(b)

Fig. 6 (a) Line scans of $\Delta R$ with a GMR sensor with a length of 100 μm and a height of 20 μm over a 50 μm wide and 0.3 μm high conducting line, which was fed with $i_s$ between 5 mA and 0.5 mA with $f$ = 1 kHz, (b) line scans of $\Delta R$ with a GMR sensor with a length of 10 μm and a height of 20 μm over the same conducting line, which was fed with $i_s$ between 25 mA and 2.5 mA with $f$ = 1 kHz.

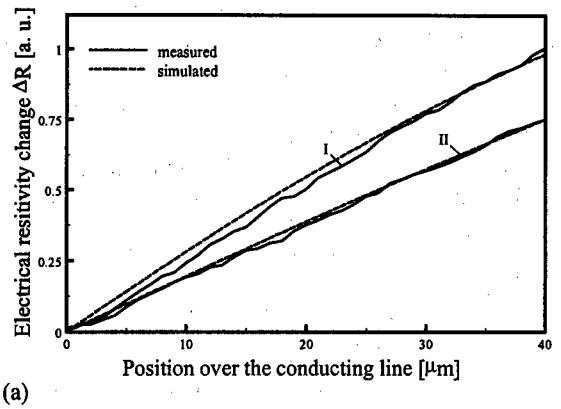
magnitude and phase can be measured. The measured magnitude and the phase of $\Delta R$ corresponds to the signal $\vec{u}_{total}$, which results from the offset $\vec{u}_{offset}$ and the real GMR sensor signal $\vec{u}_{\Delta R}$. In order to measure the magnitude and the phase of the offset the current source for the GMR sensor was inactivated during the line scan (see fig. 2.(b)). The real GMR sensor signal $\vec{u}_{\Delta R}$ for the detected current is then obtained by vectorial subtraction of the measured offset $\vec{u}_{offset}$ from $\vec{u}_{total}$. With the shielded GMR sensor currents of 10 mA with a frequency $f$ between 1 kHz and 100 kHz were detected. The obtained line scan of $\Delta R$ from the detected current with a frequency $f$ of 1 kHz shows a symmetry (see fig. 5). It is compared with the recalculated line scan of $\Delta R$ from the detected current with a frequency $f$ of 100 kHz. A very good agreement between the line scans can be achieved.

In order to investigate the possibility of the detection of currents in IC's the size of the GMR sensor was reduced by focussed ion beam (FIB). The magnetic field sensitive layer of the GMR sensor could be splitted to a desired size. Two GMR sensors with a length $l_1$ = 100 μm and $l_2$ = 10 μm and both with a height $s$ = 20 μm were realized. The results of the measurements obtained with these GMR sensors are shown in figure 6.(a) and (b). The current sensitivity decreases with reduction of length $l$ of the GMR sensor. With the GMR sensor with a length of 10 μm a current of 2.5 mA could be detected. We expect, that the current sensitivity can be improved if the height $s$ of the GMR sensor is reduced.

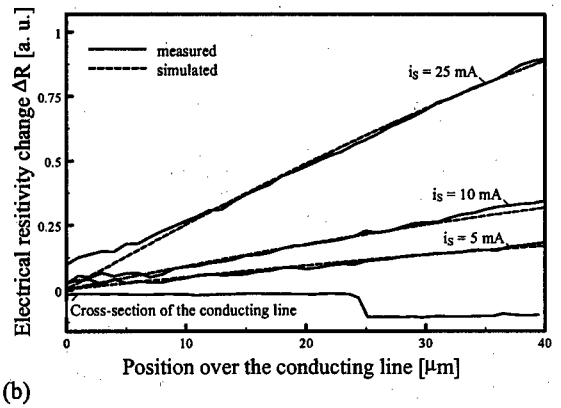
### *4.2 Quantitative current measurement*

For quantitative current measurements we used the magneto-resistive probe as shown in figure 3. Due to the mounted cantilever the used GMR sensor could not be electrically shielded.

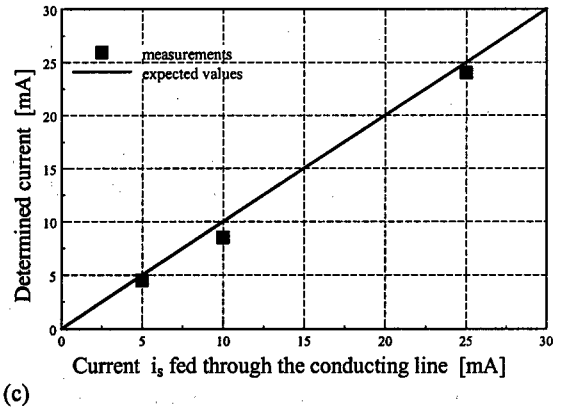
For quantitative current measurements it is essential to know the exact scan height $h$ between the GMR sensor and the surface of a device under test (DUT). By attaching a cantilever with a tip at its end under the GMR sensor we are able to place the magneto-resistive probe in a reproducible height. However, due to tolerances by the attachment of the cantilever the absolute vertical distance between the tip and the GMR sensor and therefore, the exact height $h$ is unknown. Therefore, a calibration procedure has to be developed.




(a)




(b)




(c)

Fig. 7 (a) Comparison between the obtained line scans of $\Delta R$ and simulated curves with a scan height $h$ of 58 μm : curve I) measured over a 13 μm wide and 3 μm high conducting line and curve II) measured over a 150 μm wide and 0.3 μm high conducting line, (b) determination of currents from the obtained line scans of $\Delta R$ by a fit of simulated curves, and (c) comparison between the determined currents and the currents fed through the conducting line.

The scan height $h$ can be determined from the ratio of the line scans of $\Delta R$, which are measured over two current-carrying conducting lines with different cross-sections. In order to do so, the calibration for the determination of the scan height $h$ was performed on a 150 μm wide and 0.3 μm high conducting line and a 13 μm wide and 3 μm high conducting line fed both with the same current. In the first step the topography of a conducting line was measured. Then the magneto-resistive probe was shifted of 200 μm in x-direction for measuring a line scan of $\Delta R$ (see fig. 3), because in the horizontal direction the tip is about 200 μm away from the GMR sensor. To determine the scan height $h$ the magnetic field distributions of $H_{ext,N}$ over the two conducting lines were simulated in dependence of the scan height $h$ [10], and the obtained line scans of $\Delta R$ from the detected current were fitted (see fig. 7.(a)). For the scan height $h$ of 58 μm a good agreement between the measured line scans and the simulated curves was achieved.

Experimental results of quantitative current measurements were then obtained on a 50 μm wide and 0.3 μm high conducting line, which was fed with currents $i_s$ between 25 mA and 5 mA with a frequency $f$ of 1 kHz (see fig. 7.(b)). To determine the currents fed through the conducting line the line scans of $\Delta R$ were fitted by the simulated curves. Figure 7.(c) shows the comparison between the determined currents and the currents fed through a conducting line. The result of the quantitative current measurements shows maximum measurement error of 15 %. We expect, that the accuracy of the quantitative current measurement can be improved by using a electrically shielded GMR sensor.

## 5 Conclusion

A new test method for the detection of current flow through a conducting line was realized by using a scanning magneto-resistive probe microscope. A GMR sensor was used as a magneto-resistive probe for detecting a current on a conducting line. With the GMR sensor a current of 20μA with a frequency $f$ of 1 kHz could be detected on a 4 μm wide and 0.5 μm high conducting line. To improve the detectable frequency range of currents the GMR sensor was electrically shielded. The detection of currents with frequencies up to 100 kHz was achieved.

For quantitative current measurements a new magneto-resistive probe consisting of a GMR sensor and a cantilever with a tip was developed. The quantitative current measurement with the new magneto-resistive probe was demonstrated on a 50 μm wide and 0,3 μm high conducting line by measuring currents between 25 mA and 5 mA with a frequency $f$ of 1 kHz.

## Acknowledgements

The authors would like to thank Dr W. Clemens from Siemens AG Erlangen for the kind disposal of the GMR sensors.

## References

[1] Leyk A. and Kubalek E., Electron. Lett. Vol. 31 No 24 (1995), 2089-2091

[2] Weingarten K., Rodwell M. and Bloom D., IEEE J. Quantum Electron. Vol. QE-24 No. 2 (1988) 198-220

[3] Chandrachood J., Majundar A. and Carrejo J., IEEE Electron Dev. Lett. Vol. 16 No. 7 (1995) 312-315

[4] Cramer R., Bruchaus L. and Balk L., Microelectron. Reliab. Vol. 37 No. 10/11 (1997) 1583-1586

[5] Campbell A, Cole E., Dodd B. and Anderson R.E., Microelectron. Eng. 24 (1994), 11-22.

[6] Helmreich K., Nagel P., Wolz W.and Müller-Glaser K., Microelectron. Eng. 16 (1992), 173-182.

[7] Lee T., Dantsker E. and Clarke J., Rev. Sci. Instrum. Vol. 67 No. 12 (1996) 4208-4215

[8] Tsang C., Fontana R., Lin T., Heim D., Speriosu V., Gurney B. and Williams M., IEEE Trans. Magn. Vol. 30 N0. 6 (1994) 3801-3806

[9] Clemens W., Van den Berg H.,and Rupp G., J. Appl. Phys. Vol. 81 No. 8 (1997), 4310-4312.

[10] Küpfmüller E. and Kohn E., Theoretische Elektortechnik und Elektronik (14th edn). Spinger, Berlin, pp 1993, 296-301.

PERGAMON

Microelectronics Reliability 38 (1998) 975–980

MICROELECTRONICS RELIABILITY

# Automatic Fault Tracing by Successive Circuit Extraction from CAD Layout Data with the CAD-Linked EB Test System

**K. Miura, K. Nakamae and H. Fujioka**

Department of Information Systems Engineering, Faculty of Engineering,
Osaka University
Yamada-Oka 2-1, Suita, Osaka, 565-0871 Japan

## Abstract

We describe the automatic fault tracing of a self-made 8-bit microprocessor by successive circuit extraction from CAD layout data with the CAD-linked EB test system. The origin of the marginal fault demonstrated by the shmoo plot was automatically localized. The number of probing points was 26. The total time required for fault tracing was 53 minutes. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

As LSI circuits have become denser and more complex, the performance faults such as the delay fault have been a serious issue in the diagnosis of faulty VLSI chips [1]. The CAD-linked electron-beam (EB) test system has become an indispensable instrument for probing internal behavior of VLSI circuits [2-11]. Through the fault localization process by using the method such as the guided probe diagnosis with the CAD-linked EB test system, the gate or cell that has the faulty output and all good inputs is found: a gate-level fault or a cell-level fault is determined [9, 10]. In order to search the cause of the performance faults, however, it becomes necessary to trace the fault in the lowest level of the circuits, that is, in the transistor-level [11-14].

In the CAD-linked EB test system, the node correspondence between layout and netlist in the CAD database is made. Then the alignment between the actual scanning electron microscope (SEM) image of the VLSI chip and the layout to control the movement of the X-Y stage of the EB test system is carried out. The observed point or area of the real chip is obtained by selecting a net in the netlist or a layout pattern in the layout of the CAD database. In order to link the CAD schematic or netlist and the mask layout data mutually, it is necessary to prepare the CAD data that include the data in the transistor-level. This preparation may take much time. In some situations of the CAD database system, there often arises the case where the transistor-level data is lost and only a layout data is available. For example, ASIC design does not always require the transistor-level data. In those cases, it is

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00097-3

necessary to extract the transistor-level netlist from the layout data.

To deal with the situation described above, we have developed an automatic EB fault tracing system [15] based on the hierarchical fault tracing algorithm by successive circuit extraction from CAD layout data [16]. The system is applicable to the case where only CAD layout data is available. In the method, partial circuit data along the tracing path are extracted successively on demand of the fault tracing algorithm. The hierarchical fault tracing method allows us to trace a fault hierarchically from the top-level cell to the lowest primitive cell and from the primitive cell to the transistor-level circuit in a consistent manner independent of circuit functions.

In this paper, we apply the system to a self-made 8-bit microprocessor that has a marginal fault demonstrated by the shmoo plot of an LSI tester in order to show validity of the system.

## 2. System overview

### 2.1. *Hardware description*

The hardware organization of our system is illustrated in Fig. 1. It consists of an EB test system (Schlumberger IDS 5000ZX), an LSI test system (HEWLETT PACKARD HP83000), a server computer (HEWLETT PACKARD HP9000/J210: 168 SPEC int 92, 269 SPEC fp 92) and a host computer (Silicon Graphics Indigo 2: 119 SPEC int 92, 131 SPEC fp 92) which are linked by an ethernet-based network. The appearance of this system is shown in Fig. 2.

The EB test system has an inverted, moving column and stationary DUT that permits tabletop access to device pins and direct docking to the test head of the LSI tester. Thus reliable, high-bandwidth connections are made easily, even for high pin-count devices.

The LSI test system allows us to test devices with 176 pins and 120 MHz clock frequency at present.

The server computer can treat CAD layout data for 10 chips where each chip has about 1 million transistors.

The host computer can treat a hierarchically structured CAD layout data with 1 million transistors for the fault tracing.

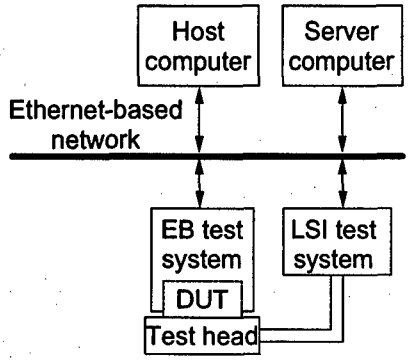



**Fig. 1** Hardware organization of the automatic fault tracing system.

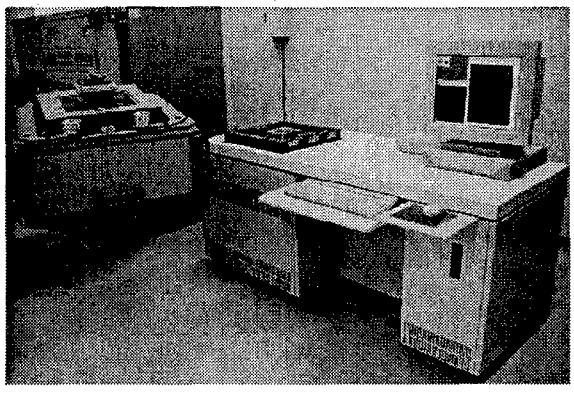


**Fig. 2** The appearance of the system.

### 2.2. *Software Description*

In order to realize the procedure in the automatic fault tracing system described, we implemented the software by integrating the following six programs:

(1) Fault tracing,

(2) Optimal probing point selection for waveform measurements,

(3) Matching of DUT interconnection pattern with CAD layout,

(4) Waveform comparison,

(5) Control of the EB tester,

(6) Control of the LSI tester.

The fault tracing program needs only CAD layout data in GDS-II format and can treat the hierarchically structured layout data. When the start point on the layout and the faulty period are given, the program outputs the upstream and observable interconnection of the start point and the period to be measured. The program is applicable to VLSI that includes sequential circuits and bi-directional busses. The preprocessing of CAD layout and the fault tracing program constitutes the main program.

The optimal probing point selection program [17] receives data from the fault tracing program. In the data, a set of polygons that constitute the interconnections around the upstream interconnection to be measured and information as to the flow of signal are included. The optimal probing point is searched by applying a set of rules. The rules are, a) to select probing points on wider electrodes, b) to select the points over the most superficial layer, c) to avoid the contact areas, d) to avoid areas where strong local field effects could negatively influence the measurement, e) to select the points on the downstream of an electrical net, f) to select the points that are not screened by other interconnections in more superficial layers. The program outputs the X and Y coordinates of the selected optimal probing point.

As the matching of SEM image with CAD layout image, we adapt a pattern matching method [18] using a hybrid genetic algorithm (GA) in which deterministic transformations are carried out on the chromosomal representation of the matching parameters. The X and Y coordinates of the top-left corner in the SEM image, the magnification of the SEM image to the CAD layout image, and the rotation angle of the SEM image to the CAD layout image are matching parameters and are outputted by the program. These matching parameters enable us to calculate the electron beam positioning error.

We adapt a waveform comparison method [19] based on the parameter extraction of the waveform since quick and precise comparison between the measured and the reference waveforms is required. Amplitude, rising edge, rise time, falling edge, fall time, and so on are waveform parameters. In the waveform parameter extraction process, two methods, the function approximation method (FAM) and the filtering method (FIL) are selectable. The FAM stands the noise that is included in the measured waveform,

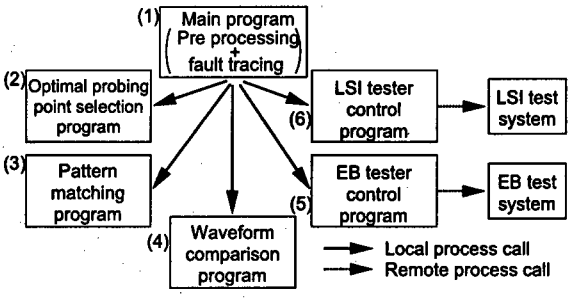



**Fig. 3** Software organization of the automatic fault tracing system.

but takes much time for the parameter extraction. Conversely, the FIL has a short processing time but is sensitive to noise. The waveform comparison program decides whether the measured waveform is good or faulty by comparing extracted parameters. In the case of the faulty signal, the faulty period is also outputted. When it is difficult to decide, the program outputs that the decision is unattainable.

The control of the EB tester is programmed by using IDS Command Line Application Programming Interface (Schlumberger Technologies, Inc.)

The control of the LSI tester is carried out by a test flow program written in C.

The software organization is shown in Fig. 3. Programs (2) to (6) are called by local process calls in the main program that include preprocessing required for each program and the fault tracing algorithm by successive circuit extraction from CAD layout data. Furthermore, programs (5) and (6) control the EB test system and the LSI test system respectively by remote process calls. The data communication between programs uses a pipe facility or a file form.

### 2.3. *System performance*

The system performance is shown in Table 1 where an acquired SEM image is a 512×512 array with 256 gray levels. The time for the main program is mainly occupied by the preprocessing time of CAD layout data and depends on the CAD layout data size. In the table, the time for the data size of 130 kB is described. The main program is called once per DUT. Thus, the total time required for the fault tracing is determined mainly by the waveform measurement.

**Table 1** System performance.

| program | content | time/call [sec] |
|---|---|---|
| main | preprocessing and fault tracing | ~140 (depends on CAD layout data size) |
| optimal probing point selection | | 10 |
| matching of DUT interconnection pattern with CAD layout | | 14 |
| waveform comparison | FAM | 17 |
| | FIL | 0.3 |
| control of the EB tester | X-Y stage movement | 5 |
| | SEM image acquisition | 10 |
| | waveform measurement | 30 |
| control of the LSI tester | | $\ll 1$ |

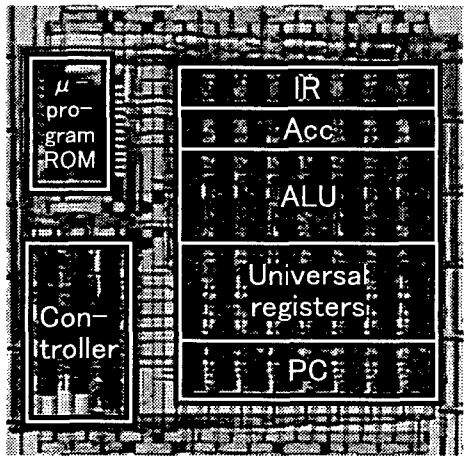



**Fig. 4** Appearance of self-made 8-bit microprocessor chip.

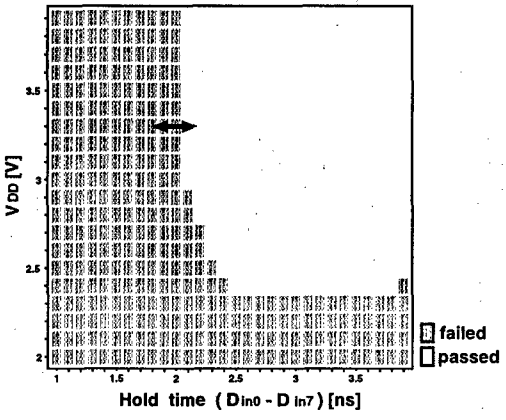



**Fig. 5** Shmoo plot for hold-time of data inputs.

## 3. Application

We applied the test system to a self-made 8-bit microprocessor LSI that contains about 4000 transistors. The processor is designed in a two metal layer and one polysilicon layer 0.5$\mu$m CMOS technology. Figure 4 shows the appearance of the chip. As is shown in the figure, the processor consists of an ALU, registers (instruction register (IR), accumulator (Acc), universal registers, and program counter (PC)), a controller, and $\mu$-program ROM. At first, the chip was tested with the LSI tester. A shmoo plot shown in Fig. 5 was obtained. In the chart, the X axis is the hold-time $t_h$ of data inputs $D_{in0}$–$D_{in7}$, and the Y axis is the power-supply voltage $V_{DD}$: the boundary between the good state and the faulty state lies around $t_h$ = 2.0ns. We traced the marginal fault indicated by the shmoo plot.

An automatically executed result is shown in Table 2. Each line of the table shows the sequence of probing, interconnection number, measured period, comparison result, and faulty signal period. Some other important comments are also described. Rows from 2 to 23 in Table 2 show the fault tracing results in the cell level: the faulty cell is specified. In order to explain the situation of fault tracing, we show the cell level circuit translated from successively extracted circuit data in Fig. 6, where the fault tracing path is drawn by thick solid lines. The cell DATAPATH7 lightly crosshatched in Fig. 6 is specified. Then the fault tracing proceeds to the next lower level. In this level, the cell DFF heavily crosshatched in Fig. 6 is specified. Rows from 24 to 29 show the

**Table 2** Executed result.

```
 1: Initial fault phase: [100, 130]
 2: measure[ 1]:142, Phase[0, 120], Faulty[100, 120]
 3: measure[ 2]:32, Phase[0, 120], Good
 4: measure[ 3]:41, Phase[0, 120], Good
 5: measure[ 4]:44, Phase[0, 120], Good
 6: measure[ 5]:46, Phase[0, 120], Good
 7: measure[ 6]:51, Phase[0, 120], Good
 8: measure[ 7]:53, Phase[0, 120], Good
 9: measure[ 8]:43, Phase[0, 120], Good
10: measure[ 9]:61, Phase[0, 120], Good
11: measure[10]:24, Phase[0, 120], Good
12: measure[11]:71, Phase[0, 120], Good
13: measure[12]:25, Phase[0, 120], Good
14: measure[13]:96, Phase[0, 120], Good
15: measure[14]:121, Phase[0, 120], Good
16: measure[15]:91, Phase[0, 120], Good
17: measure[16]:73, Phase[0, 120], Good
18: measure[17]:89, Phase[0, 120], Good
19: measure[18]:92, Phase[0, 120], Good
20: measure[19]:30, Phase[0, 120], Good
21: measure[20]:75, Phase[0, 120], Good
22: There is a fault in the cell: DATAPATH7
23: There is a fault in the cell: DFF
24: measure[21]:894, Phase[0, 120], Faulty[100, 120]
25: measure[22]:893, Phase[0, 120], Faulty[100, 120]
26: measure[23]:899, Phase[0, 120], Good
27: measure[24]:897, Phase[0, 120], Faulty[100, 120]
28: measure[25]:895, Phase[0, 120], Good
29: measure[26]:898, Phase[0, 120], Good
30: There may be a fault around the interconnection 897
31: interconnection(897)
32: MOS-FET(4751)
33: MOS-FET(4750)
34: MOS-FET(4735)
35: interconnection(913)
36: MOS-FET(4734)
37: MOS-FET(4733)
38: interconnection(918)
39: MOS-FET(4747)
40: MOS-FET(4746)
41: MOS-FET(4731)
42: MOS-FET(4730)
43: interconnection(922)
44: MOS-FET(4745)
45: MOS-FET(4729)
46: interconnection(976)
47: MOS-FET(4749)
```

fault tracing results in the transistor level. Figure 7 is the transistor-level circuit drawn by successively extracted circuit data. Rows from 30 to 47 suggest the name of possible faulty interconnections and devices, that are crosshatched in Fig. 7.

As is described above, the origin of the fault is localized automatically. The number of probing points was 26. The total time required for fault tracing was 53 minutes.

## Conclusions

We described the automatic fault tracing of a self-made 8-bit microprocessor by successive circuit extraction from CAD layout data with the CAD-linked

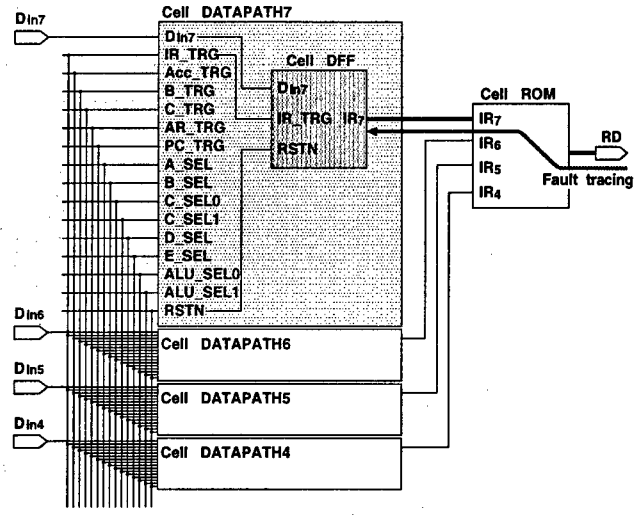



**Fig. 6** Cell level tracing.

EB test system. The origin of the marginal fault demonstrated by the shmoo plot was automatically localized. The number of probing points was 26. The total time required for fault tracing was 53 minutes.

## Acknowledgments

We would like to express our appreciation to Schlumberger Technologies, Inc. for their cooperation.

The VLSI chip in this study has been fabricated in the chip fabrication program of VLSI Design and Education Center (VDEC): the University of Tokyo with the collaboration by NTT Electronics Corporation and Dai Nippon Printing Corporation.

This research was supported by a Grant-In-Aid for Scientific Research of the Ministry of Education, Science and Culture, Japan.

## References

[1] Director S W, Maly W and Strojwas A J. VLSI design for manufacturing: yield enhancement. Kluwer Academic Publishers, Boston, 1990.

[2] Tamama T and Kuji N. Automated fault diagnosis EB tester and its application to a 40k-gate VLSI circuit. in Proc. Int. Test Conf. 1985, pp 643–649.

[3] Bergher L, Laurent J, Courtois B and Collin J P. Towards automatic failure analysis of complex ICs through e-beam testing. in Proc. Int. Test Conf. 1986, pp 465–471.

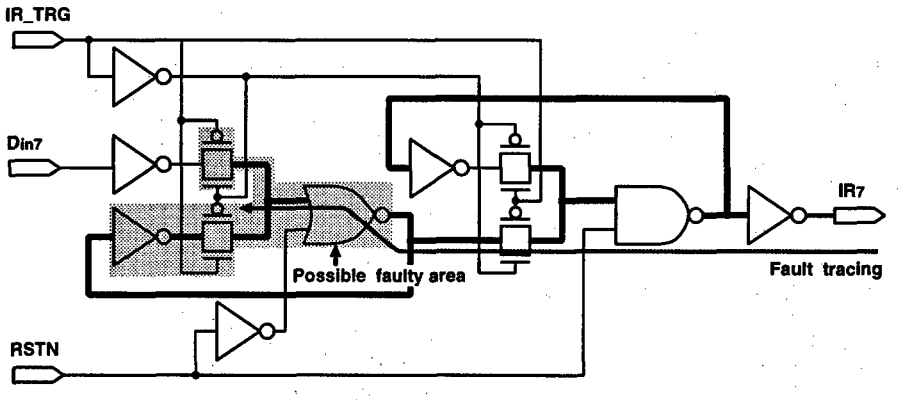



**Fig. 7** Transistor level tracing.

[4] Consina S, Liu G, Lattanzi L. Reyfaman S and Richardson N. Software integration in a workstation-based e-beam tester. in Proc. Int. Test Conf. 1986, pp 644–649.

[5] Guiguet I, Marzouki M and Courtois B. An integrated debugging system based on e-beam test. Microelectronic Engineering. 1987, vol 7, no 2–4, pp 275–282.

[6] Melgara M, Battù M, Garino P, Dowe J and Marzouki M. Fully automatic VLSI diagnosis in a CAD-linked e-beam probing system. Microelectronic Engineering. 1987, vol 7, no 2–4, pp 283–293.

[7] Ura K and Fujioka H. Electron beam testing. In: Hawkes P W (Ed.) Advances in Electronics and Electron Physics. Academic Press, New York, 1989, vol 73, pp 223–317.

[8] Hu A and Nijima H. New approach to integrate LSI design databases with e-beam tester. in Proc. Int. Test Conf.1990, pp 1040–1048.

[9] Noble A C. IDA: a tool for computer-aided failure analysis. in Proc. Int. Test Conf. 1992, pp 848–853.

[10] Yamaguchi N, Sakamoto T, Nishioka H, Majima T, Satou T, Shinada H, Todokoro H and Yamada O. E-beam fault diagnosis system for logic VLSIs. Microelectronic Engineering. 1992, vol 16, no 1–4, pp 121–128.

[11] Fujioka H, Nakamae K. New trends on beam testing of integrated circuits. in Proc. ESREF. 1993, pp 505–512.

[12] Fujioka H and Nakamae K. LSI failure analysis with CAD-linked electron beam test system and its cost evaluation. IEICE Trans. Electron. 1994, vol E77–C, no 4, pp 535–545.

[13] Miura K, Nakamae K and Fujioka H. Automatic tracing of transistor-level performance faults with CAD-linked electron beam test system. IEICE Trans. Fundamentals. 1994, vol E77–A, no 3, pp 539–545.

[14] Nakamae K, Miura K and Fujioka H. VLSI testing with CAD-linked electron beam test system. Microelectronic Engineering. 1995, vol 31 , no 1–4, pp 319–330.

[15] Miura K, Nakata K, Nakamae K and Fujioka H. Automatic EB fault tracing system by successive circuit extraction from VLSI CAD layout data. in Proc. 6th Asian Test Symposium. 1997, pp 162-167.

[16] Miura K, Nakamae K and Fujioka H. Hierarchical VLSI fault tracing by successive circuit extraction from CAD layout data in the CAD-linked EB test system. Journal of Electronic Testing: Theory and Applications. 1997, vol 10, no 3, pp 255-269.

[17] Nakagaki R, Matsumoto M, Nakamae K and Fujioka H. Automatic selection algorithm of optimal probing points for waveform measurements in CAD-linked electron beam test system (in Japanese). in Proc. Symposium on LSI Testing. Osaka, 1994, pp 123–128.

[18] Fukutome K, Nakamae K and Fujioka H. Matching of DUT interconnection pattern with CAD layout in CAD-linked electron beam test system by using a hybrid genetic algorithm (in Japanese). in Proc. Symposium on LSI Testing. Osaka, 1996, pp 24–29.

[19] Nishi T, Nakamae K and Fujioka H. Waveform comparison system using parameters extracted from signal waveforms in electron beam test system (in Japanese). in Proc. Symposium on LSI Testing. Osaka, 1995, pp 126–131.

PERGAMON

Microelectronics Reliability 38 (1998) 981–986

MICROELECTRONICS RELIABILITY

# Cantilever influence suppression of contactless IC-testing by electric force microscopy

V.Wittpahl, U.Behnke, B.Wand and W.Mertin

*Werkstoffe der Elektrotechnik, Sonderforschungsbereich 254,*
*Gerhard-Mercator-Universität Duisburg, Bismarckstraße 81, 47048 Duisburg, Germany*

**Abstract**

The Electric Force Microscopy (EFM) has been developed as a contactless test technique for chip internal analysis. The physical principle is explained by a simple model, which is proven to be insufficient for EFM based IC-testing. An extended model, which considers also the forces along the probe's cantilever, is presented. A new method based on the extended model separates the electric force at the probe's tip end from the detected force via suppression of the cantilever influence. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Many different Scanning Force Microscope applications for the detection of surface related informations have been developed, since the Scanning Force Microscope has been introduced [1]. The EFM is an application of the Scanning Force Microscope, which uses an electric biased Scanning Force Probe (SFP) to determine electric informations of the specimen via detected Electric Force Interaction (EFI). The EFI is measured by the cantilever deflection.

The test of Integrated Circuits by EFI has also been shown for time resolved digital signals [2] and frequencies up to the GHz range [3]. Time resolved measurements of fast periodical digital patters have been shown up to 3.2 GBit/s [4]. The highest frequency, that is detected as an EFI is reported by 110 GHz [5].

While the principle use of the EFM has been shown, some observed phenomena are not investigated and cannot be explained with the current capacitance model like

- unsymmetric increasing and decreasing of signals along a contact or line
- detection of high signals above an insulator surface between two conducting lines

We are going to present an extended description of the EFI for a SFP, which includes the electric force distribution along the whole SFP and the coupling with the mechanical physics of the cantilever by the deflection curve bending line.

The presented description of the EFI is used to develope a measurement procedure for the Cantilever Influence Suppression (CIS).

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00118-8

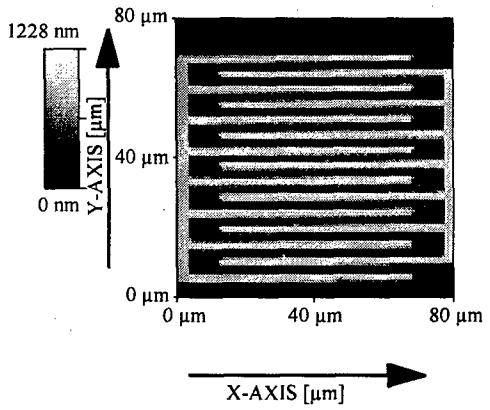



Fig. 1. Scanning Force Microscope image of the interdigital structure topography; the width of the fingers is 1.5µm

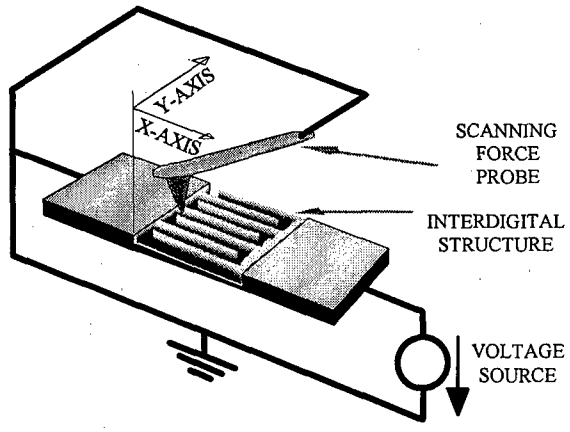



Fig. 2. Measurement set-up for the detection of the EFI

## 2. The current EFI-model

### *2.1. Theory*

The EFI is described by the electric force $F$

$$F = \frac{1}{2} \cdot \frac{dC_0}{dh} U^2 \qquad (1)$$

where $F$ is the effective force, $dC_0/dh$ the change of the tip end-specimen capacitance by variation of the distance $h$ and $U$ the applied voltage between SFP and specimen [2, 3].

The electric force $F$ causes a deflection $\Delta h$ of the SFP. The deflection is given by the Hook's Law :

$$F = K \cdot \Delta h \qquad (2)$$

where $K$ is the spring constant of the cantilever [2, 3].

### *2.2. Limits of the current EFI-model*

The current description of the EFI reaches its limit sudden, if a multiple conducting line structure, like a bus structure, is scanned. An interdigital structure (see Fig. 1) has been used to show the weak points of the EFI-model.

The SFP is grounded and positioned perpendicular to the conducting lines of the test structure (see Fig. 2). The interdigital structure is biased on the right electrode with an electric voltage at a frequency $f$.

The detected EFI is given in Fig. 3. The linescans in x- and y-direction (Fig. 3.b, Fig. 3.c) are taken from the areascan (Fig 3.a).

Typical phenomena are observed:

- unsymmetric increasing and decreasing of signals along a contact or line

The linescan $Y_2$ has been measured along a biased finger in x-direction (see Fig. 3.b). The detected EFI along the finger is not constant but increases in direction of the biased electrode. The linescan $X_3$ has been measured perpendicular to the fingers in y-direction (see Fig. 3.c). The detected EFI shown different values for the biased fingers.

- detection of high signals above an insulator surface between two conducting lines

The linescan $Y_1$ has been measured along the gap between the fingers in x-direction (see Fig. 3.b). The detected EFI is not zero but increases in direction of the biased electrode in a similar slope like the linescan $Y_2$ along the biased finger.
The linescan $X_3$ (see Fig. 3.c) shows very high EFI between the biased fingers. Furthermore, the detected EFI in y-direction above the grounded electrode (see Fig. 3.c, $X_1$) is 10% of the highest detected EFI value, while the EFI outside the interdigital structure is up to 40% (see Fig. 3.c, $X_3$).

These presented phenomena cannot be explained by the current EFI model.

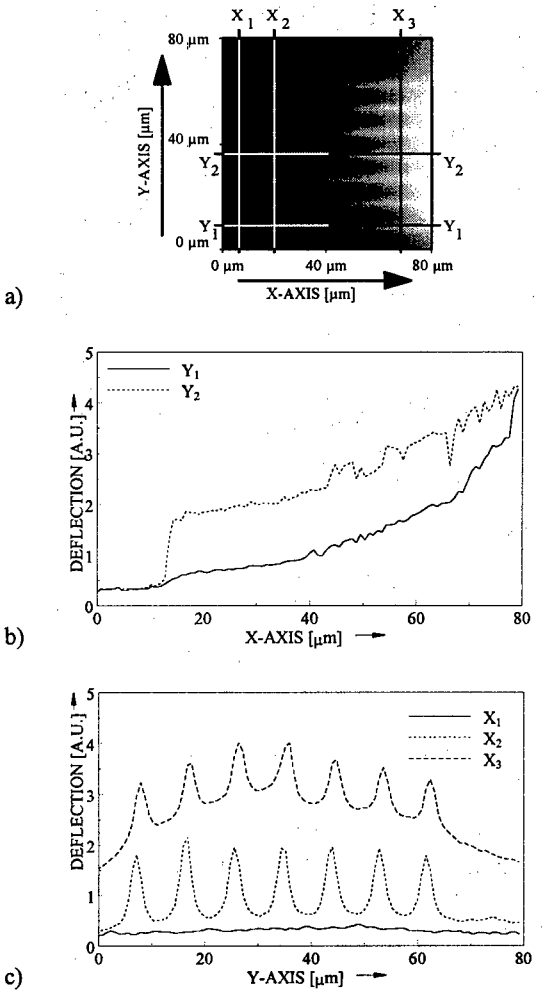



Fig. 3. Measured EFI scan arcoss the interdigital structure; work frequency 10 GHz
a) Areascan
b) Selected x-direction linescans taken from a)
c) Selected y-direction linescans taken from a)

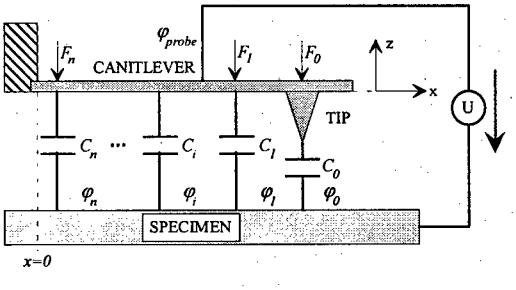



Fig 4. Simplified schematic of the extended EFI-model; the SFP represents an net node in an electric equivalient circuit for IC-testing use

## 3. The extended EFI-model

### 3.1. Theory

#### 3.1.1. Extention of the electric description

Every conducting line below the SFP builds a capacitance with the biased SFP, when the EFM is used for IC-testing. The SFP represents now in the electronic circuit for the IC-testing set-up a net node of the potential $\varphi_{probe}$ with a set of capacitances (see Fig. 4.). Every conducting point $i(x_i, y_i)$ on the specimen surface with the potential $\varphi_i(x_i, y_i)$ builds a small capacitance $C_i(x_i, y_i)$ with the corresponding point on the SFP in the height $h_i(x_i, y_i)$. The force $F_i$ at point $i$ on the cantilever surface can be described by

$$F_i = \frac{1}{2} \cdot \frac{dC_i}{dh} \cdot (\varphi_{probe} - \varphi_i)^2 \tag{3}$$

where $dC_i/dh$ is the change of the capacitance $C_i(x_i, y_i)$ by variation of the distance $h_i(x_i, y_i)$. The applied voltage is given by the potential difference of the SFP potential $\varphi_{probe}$ and the specimen potenial $\varphi_i(x_i, y_i)$ at the point $i(x_i, y_i)$. The resulting force $F$ along the whole SFP can be given as the sum of the forces of the tip $F_{tip}$ and of the cantilever $F_{cant.}$

$$\begin{aligned} F &= F_{tip} + F_{cant.} \\ &= \frac{1}{2} \cdot \frac{dC_0}{dh} \cdot (\varphi_{probe} - \varphi_0)^2 \\ &+ \frac{1}{2} \cdot \sum_{i=1}^{n} \frac{dC_i}{dh} \cdot (\varphi_{probe} - \varphi_i)^2 \\ &= \frac{1}{2} \cdot \sum_{i=0}^{n} \frac{dC_i}{dh} \cdot (\varphi_{probe} - \varphi_i)^2 \end{aligned} \tag{4}$$

#### 3.1.2. Extention of the mechanic description

The exact form of the deflected cantilever in relation to different affecting forces along the cantilever is described in mechanics by the deflection curve bending line. The probe is simplified by an one side anchored cantilever of the length $L$.

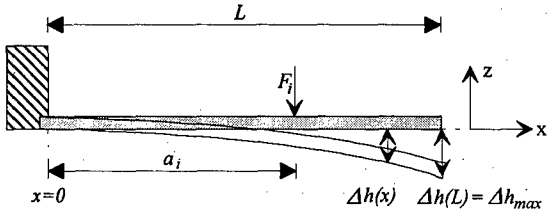



Fig. 5. Principle of the deflection curve bending line

The size of the tip is neglectable in our consideration for the mechanic description of the bending line, because the tip is much smaller than the cantilever. So the mechanic model is simplified like in Fig. 5. If a force $F_i$ is acting at the point $a_i$, the deflection $\Delta h$ for each point x along the cantilever is given by [6]

$$\Delta h(x) = \frac{F_i \cdot x^2}{6 \cdot E \cdot I} \cdot \left(a_i - \frac{x}{3}\right) \qquad (5)$$

where $E$ is the modulus of elasticity and $I$ is the area moment of intertia.

### 3.2. Simulation

#### 3.2.1. Simulation program

A simulation has been written that allows the calculation of the cantilever deflection caused by EFI for a linescan above a mutiple conductuing line structure. The simulation enables the calculation of the total deflection or just the deflection that is caused by forces on a restricted area of the SFP.

Images of a used SFP taken by a Scanning Electron Microscope are shown in Fig. 6.a). The cantilever is simulated as a bar with the length of 480μm, the width of 60μm and the thickness of 4μm. The tip is simulated by a pyramide with the measured shape from the images of Fig. 6.a). The pyramide is positioned 35μm from the cantilever end. The cantilever end is fitted as a triangle (see Fig 6.b).

#### 3.2.2. Simulation results

To demonstrate the cantilever influence for a simple test structure, the EFI-linescans are measured and simulated (see Fig. 7.). The test arrangement is shown in Fig. 7.a). The SFP is grounded and is scanned perpendicular to three conducting lines. The distance between the lines is 10μm. The outer lines expand infinite in relation to the SFP.

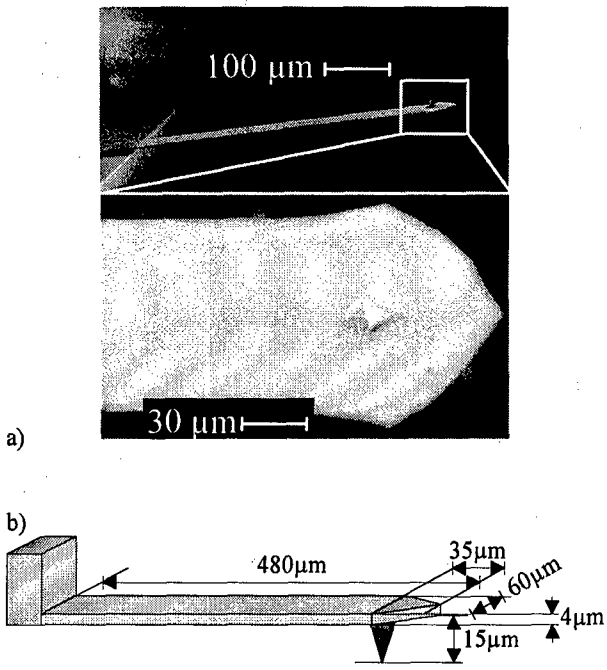



Fig. 6. Typical SFP
a) Images of a SFP taken by a Scanning Electron Microscope
b) Simulated SFP

The middle conducting line with a width of 25μm is biased with 1 Volt while the other lines are grounded.

The linescans are taken with a tip end-specimen surface distance of $h_1 = 125\text{nm}$ and $h_2 = 185\text{nm}$. The measured and the simulated linescans are shown is Fig. 7.b). The diagram includes furthermore a linescan of the topography and the expected linescan of the current EFI-model.

While the current EFI-model predicts only an EFI above the electric biased line, the SFP is detecting also a signal above the grounded lines of up to 40% of the maximum EFI on the biased line. The EFI linescan shows also an unsymmetric slope above the grounded lines. The simulations and the measurements show in principle the same behaviour. The correlation is very good above the middle conducting line. There is a difference between the curves above the grounded lines. This difference is due to the unknown exact form of the SFP. It is also obvious from Fig. 7.b) that the increaseing and decreaseing edges have the same values for both tip end-specimen surface distances.

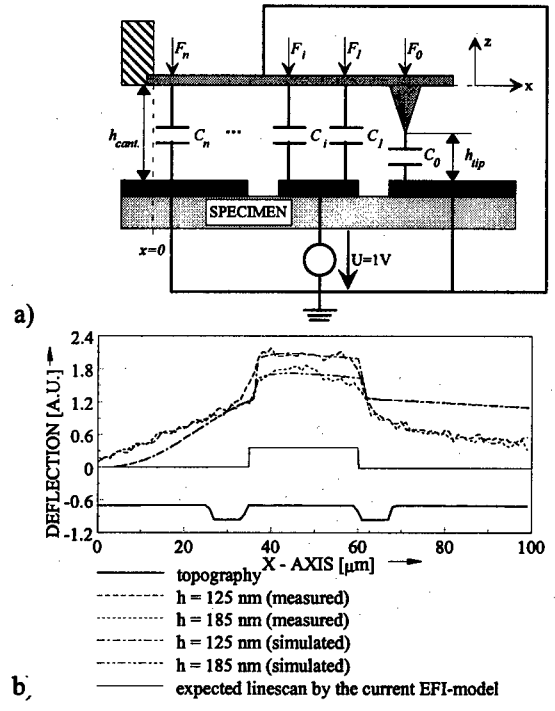



Fig. 7.
a) Set-up of the simulated linescan; the SFP is scanning perpendicular to the conducting lines.
b) EFI linescans

## 4. The Cantilever Influence Suppression

### *4.1. Principle*

The cantilever-specimen surface distance $h_{cant.}$ is approximate the tip height of 10µm to 15µm. The typical tip end-specimen surface distance $h_{tip}$ is 100nm to 500nm. Therefore, the relation $h_{tip} : h_{cant.}$ is around 1:100. When the distance is varied of about $\Delta h \approx 100\,\text{nm}$, $h_{tip}$ is changed up to 100% while $h_{cant.}$ is approximate constant in respect to $h_{tip}$ and $h_{cant.}$.

The change of the capacitance for the cantilever-specimen surface $C_{cant.}$ and the tip end-specimen surface $C_{tip}$ in relation to the tip end-specimen surface distance $h_{tip}$ is plotted in Fig. 8. The curves are normalized by their start value at $h = 20\text{nm}$. While the capacitance $C_{cant.}$ is approximately constant for the distance variation of 20nm to 500nm, the capacitance $C_{tip}$ decreases

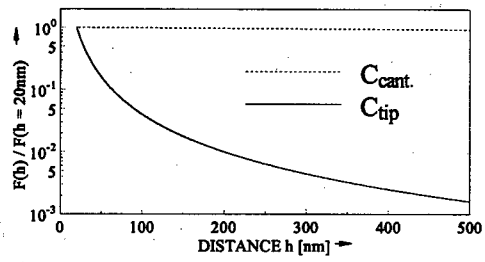



Fig. 8. Relation between the force at the capacitance of the cantilever and the tip end in relation to the SFP-surface distance. The curves are normalized by the start value at 20nm

over two decades. It enables the possibilty to eliminate the cantilever influence from the measurement result. This measurement procedure is called CIS.

If the EFI is detected in two different distance $h_1$ and $h_2 \approx h_1 + \Delta h$ and the detected forces are substracted, the force difference $\Delta F$ is

$$\begin{aligned} \Delta F &= F(h_1) - F(h_2) \\ &= (F_{tip}(h_1) + F_{cant.}(h_1)) \\ &\quad - (F_{tip}(h_2) + F_{cant.}(h_2)) \end{aligned} \tag{6}$$

If the assumption $F_{cant.}(h_1) \approx F_{cant.}(h_2)$ is used, the force difference $\Delta F$ becomes

$$\begin{aligned} \Delta F &= F_{tip}(h_1) - F_{tip}(h_2) \\ &= \frac{1}{2} \cdot \left( \left.\frac{dC_0}{dh}\right|_{h=h_1} - \left.\frac{dC_0}{dh}\right|_{h=h_2} \right) \cdot (\varphi_{probe} - \varphi_0)^2 \end{aligned} \tag{7}$$

The force difference $\Delta F$ depends only on the capacitance $C_{tip}$.

### *4.2. Measurements*

To proof the principle of the CIS, the EFI-linescans of Fig. 7 are subtracted and the force difference $\Delta F$ is shown in Fig. 9. The simulated linescans show a force difference $\Delta F$ of zero outside the electric biased line and a constant signal across the electric biased line. This is exact the result that is expected by the current EFI-model. The measured force difference $\Delta F$ shows a background noise that causes an EFI close to zero outside the electric biased line and a variation of the signal along the electric biased line. The results in Fig. 9.

demonstrate that the CIS is working. The cantilever influence is neglegable on the force difference $\Delta F$ for the simulated EFI-linescans as well as for the measured EFI-linescans.

The CIS is used now for the EFI-scan on the interdigital structure of Fig. 3. The force difference $\Delta F$ is taken for $h_1 \approx 60\text{nm}$ and $h_2 \approx 350\text{nm}$. The result of the CIS measurement is presented in Fig. 10. The increase of the EFI in x-direction has disappeared. The linescan $Y_2$ in Fig. 10.b) shows a constant EFI signal with a background noise offset along the biased finger. The EFI is zero for $Y_2$ outside the biased finger as well as for $Y_1$ between the fingers and above the grounded electrode. The cantilever influence disappears also in y-direction. The EFI is zero arcross the grounded fingers as well as outside the interdigital structure (see Fig. 10.c).

## 6. Conclusion

It has been shown that the current EFI-model for a SFP is not complete. An extended model that includes the electric and mechanical physics of the cantilever has been introduced. The correctness of this model has been shown by simulations and measurements. Finally a method is presented that enables a CIS for EFI-scans with a SFP.

## Acknowledgements

The authors are very grateful to the Deutsche Forschungsgemeinschaft for the financial support of the project.

## References

[1] Binnig G., Quate C.F. and Gerber Ch. Physical Review Letters V56No.9 (1986); 930-933

[2] Brigdes G.E., Noruttun D., Said R.A., Thomson D.J, Lam T. and Qi R. J.Vac.Sci.Technol.A 16(2) March/April (1998), 830-833

[3] Hou A.S., Nechay B.A., Ho F. and Bloom D.M. Optical and Quantum Electronics 28 (1996); 819-841

[4] Sprengepiel J., Böhm C. and Kubalek E. Microelectron. Eng. 31 (1996), 181-186

[5] Leyk A. and Kubalek E. Electronics Letters Vol. 34 No. 2 (1998), 196-197

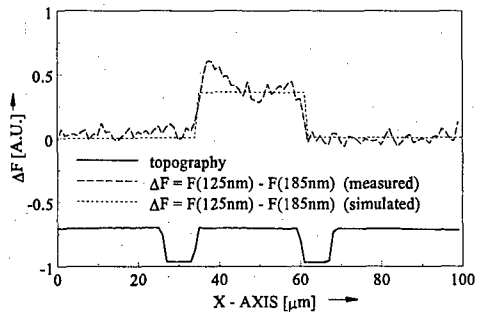



Fig. 9. CIS result of the EFI linescans of Fig. 8.

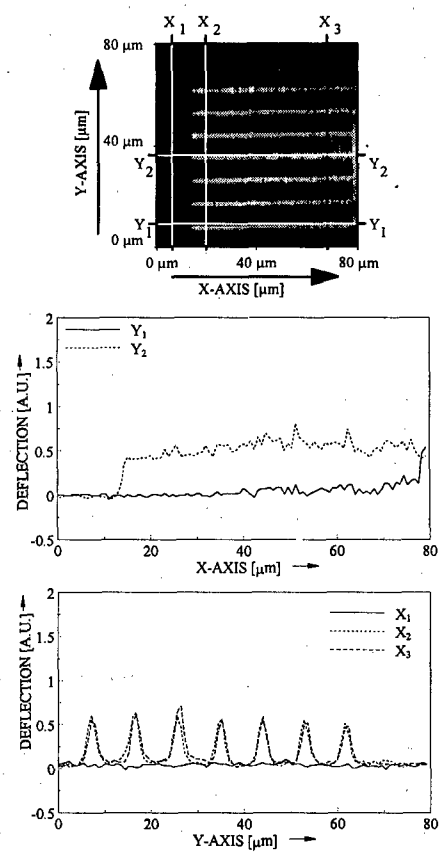



Fig. 10. CIS result of EFI scan arcoss the interdigital structure of Fig. 3 with; work frequency 10 GHz ($h_1 \approx 60\text{nm}$ and $h_2 \approx 350\text{nm}$)

a) Areascan
b) Selected x-direction linescans of a)
c) Selected y-direction linescans of a)

[6] Sarid D. "Scanning Force Microscopy", Oxford Press, Oxford, 1991

PERGAMON

Microelectronics Reliability 38 (1998) 987–992

MICROELECTRONICS RELIABILITY

# Application of Layout Overlay for Failure Analysis

C. Burmer, S. Görlich, S. Pauthner

*Siemens AG, Semiconductors, Failure Analysis, HL FA,*
*POB 801709, D-81617 Munich, Balanstr.73, D-81541 Munich, Germany*
*Tel.: +49 89 636 24572, Fax.: +49 89 636 28334, e-Mail: siegfried.goerlich@hl.siemens.de*

**Abstract**

New layout overlay techniques have been developed based on standard image correlation techniques to support failure analysis in modern microelectronic devices, which are critical to analyze because they are realized in new technologies using sub-µm design rules, chemical mechanical polishing techniques CMP and auto-routing design techniques. As the new techniques are realized as an extension of a standard CAD-navigation software using standard image format "TIFF", which is available at all modern FA-equipment, these techniques can be used for all modern failure analysis methods. Examples of application are given for circuit modification using Focused Ion Beam (FIB), for supporting preparation from the backside and for fault localization using emission microscopy. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Modern microelectronics is driven by dynamic changes in both technology and design.. Important technology improvements are:

- shrinking of design rules (sub-µm) and thereby of IC-internal dimensions;
- more conduction levels, 3D-integration;
- topography is reduced in new technologies by flatten and polishing (reflow-glass, wafer-polishing,..)

Changes in design style include:

- use of higher languages, synthesis and auto-routing of layout;
- reuse, generation and shrinking of cores, blocks or makros;

Many of these improvements for realization of more advanced circuits have unfortunately drawbacks for failure analysis. Especially the following features causes higher complexity in failure analysis, namely in chip-internal navigation:

- real sub-micron devices prevent the use of low price optical microscopes for navigation in the circuit, but force the enhanced application of highly expensive tools (SEM, FIB, EBP etc.);
- multi level wiring in combination with CMP-technology reduces observability of internal nodes and topography contrast, as every wiring level is flattened by the following isolation layer;
- synthesis and routing tools lead to a randomized look of the logic instead of a well structured block layout.

Whereas in older IC-technologies the image is providing enough surface topography for locating patterns, this is no longer true for modern technologies. This is demonstrated for FIB-images in Fig. 1. In part a) the topography is good enough to see two levels of metal, whereas in part b) the surface is totally flat as none of the underlying wiring metal exhibits any topography contrast. During the

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00075-4

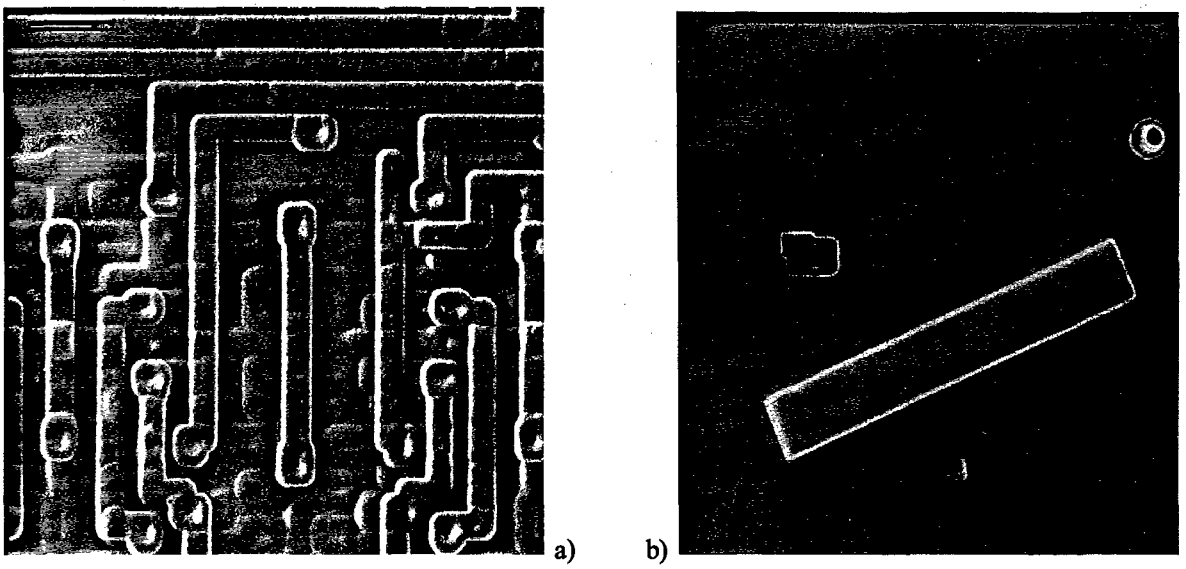



**Fig. 1**: Comparison of IC-surfaces of different technologies as observed in FIB: without CMP (a) metal layers can be used for navigation, whereas with CMP (b) this is not possible.

FIB-imaging charging may cause shifts of the image during sputtering (see squares at the left) or discharging may cause holes in the oxide layer (see upper right), which even may destroy the circuit. This complicates the circuit modification procedure.

## 2. Principle of Layout Overlay Technique

Beside specific software developed by FA-equipment houses (e.g. FIB, EBP companies [1, 2, 3]) a general approach for layout overlay to support failure analysis is demonstrated here. It can be applied to all image generating equipment like SEM, FIB, SXM, SLM, emission microscopes etc. As the image-layout overlay is realized on the side of the CAD-navigation software and not on the equipment side, and as it is based on the standard image format (TIFF), which is available at all FA-equipments, it can be used for all types of analysis at all kinds of methods. Especially it can be installed as an extension of the standard CAD-navigation software [4].

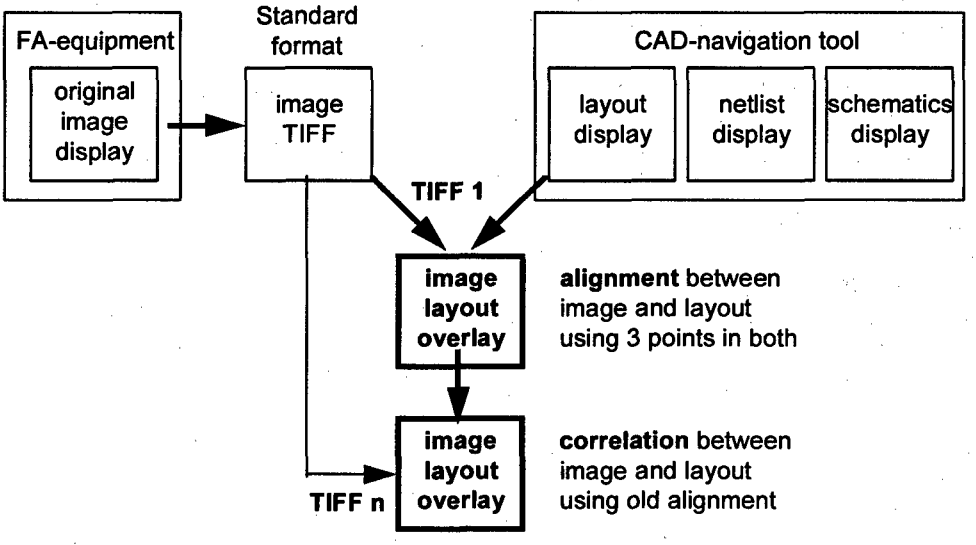



**Fig. 2**: Principle flow of layout-overlay

### 2.1 Principle Procedure

Basic feature is a 3-point alignment of a part of the layout (generated from GDS-input) viewed at the workstation with the image (TIFF-file) enabling zooming, rotating, shifting etc. This alignment (compare fig. 2) can be realized for each image, or it can be saved and used for the following images. This function is especially used in emission microscopy, when the layout is correlated to the optical (reflection) image (TIFF 1) firstly, to succeed in overlaying the following emissions image (TIFF n) with the layout data.

### 2.2 User Interface

The new function is realized as an additional function in the Maskview program via the "Image" command. [5].

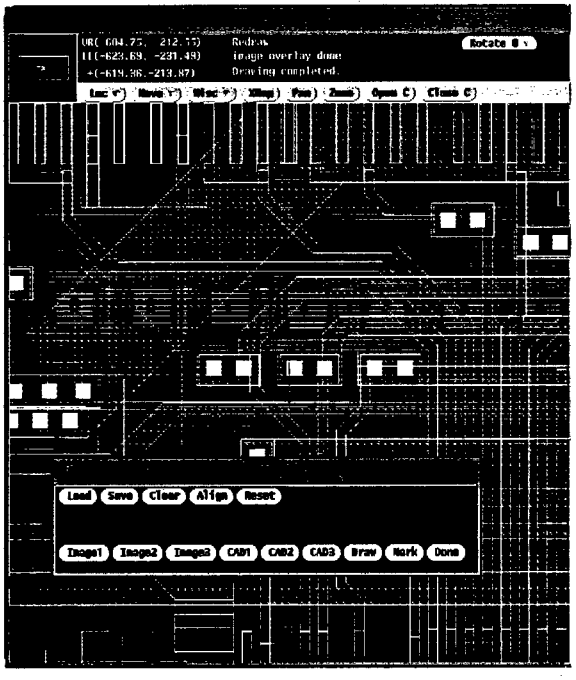



**Fig. 3**: User interface for alignment of TIFF-image to layout using 3 points in the layout-tool

In fig. 3 the status is shown after loading an specific TIFF-image, which is not yet aligned to the layout. Using the alignment menu three clearly defined, striking points well visible in both layout and image, being near the very border of the image, (e.g. corners of conduction tracks) should be chosen as alignment points "Image1" "Image2", and "Image3" and "CAD1", "CAD2", and "CAD3". This may be repeated till the accuracy of alignment is perfect. Depending on the skill of the user this procedure may take 3 -5 minutes.

## 3. Application

Examples for the realized alignment and the use of it are discussed in this paragraph. Applications are chosen from three areas:

1. preparation of chip from backside using infrared light imaging;
2. failure analysis using emission microscopy;
3. circuit modification using FIB at chip produced in CMP-technologies.

### 3.1 IR-Imaging for Backside Preparation

Multi level wiring and of modern packages, like flip chip technology, complicates the failure analysis. It is no longer possible to investigate from the top of the chip through several metal layers down to the active layers of the substrate. Therefore analyses may be done from the backside [6, 7], as silicon is transparent above 1.2 μm due to its band gap. However, the lateral resolution is reduced when using infrared light, and heavily doped substrates reduce transmission and therefore have to be thinned down to 50 μm. To support navigation at sub-μm device during backside preparation and emission microscopy again layout overlay is applied.

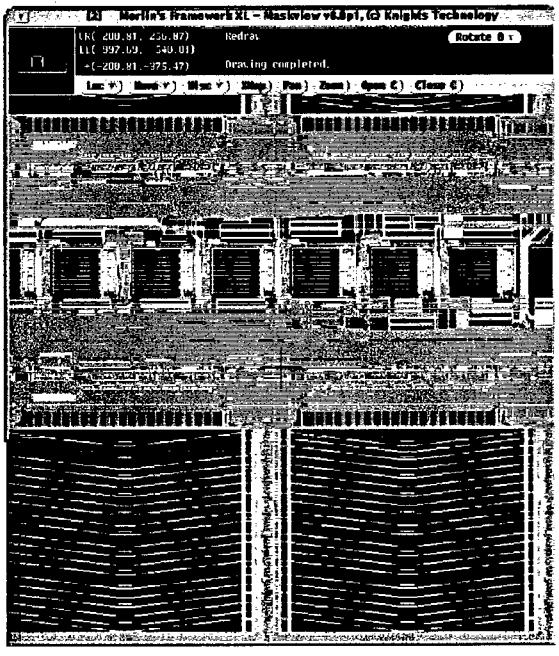
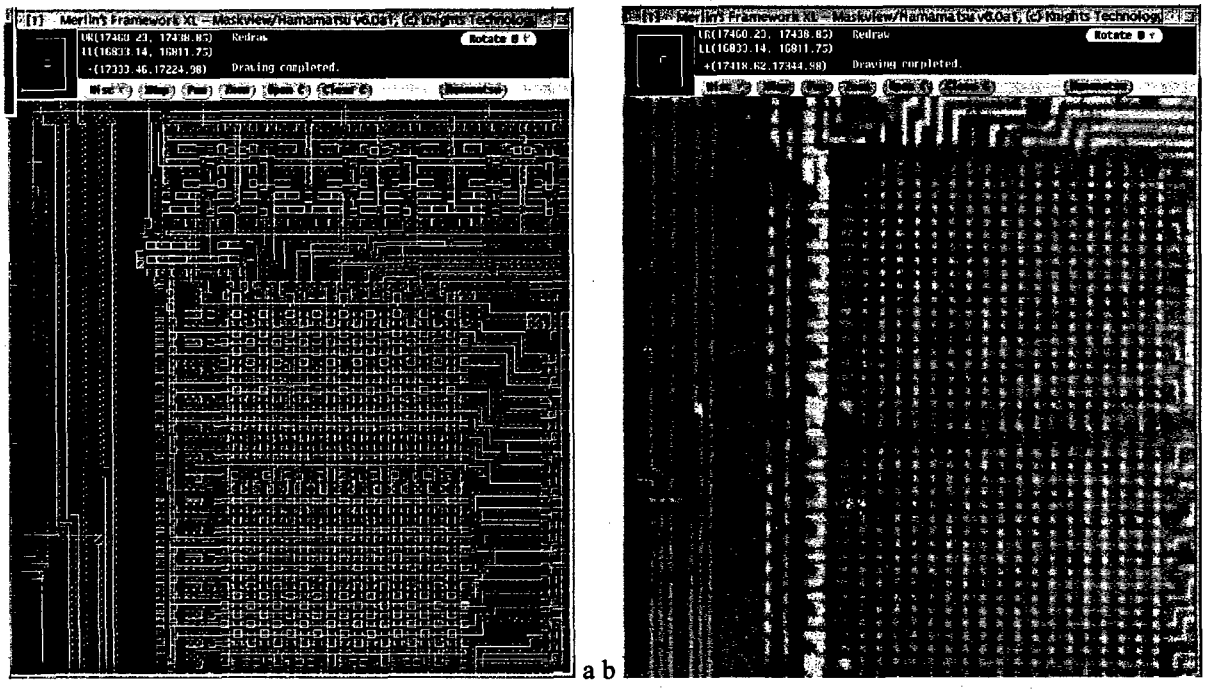



**Fig. 4**: Application in backside IR-microscopy



a b

**Fig. 5**: Layout (a) and optical image(b) of part of a micro-controller investigated by emission microscopy

Here the low resolution IR-image is aligned to the (sub-μm) layout of the CAD-navigation tool using wide conduction tracks or other well visible structures for alignment. High accurate preparation can be done locally after that alignment using the layout plan instead of the low resolution IR-Image. Fig. 4 demonstrates this overlay of an IR-image taken from the backside of an 16 Mb-SDRAM.

### *3.2 Failure Analysis using Emission Microscopy*

In the second example for this image overlay technique an emission spot is shown as it is correlated inside the layout. The electrical failure in this case was high power consumption in power-down mode of a 16-bit micro-controller. The optical reflection image was used as TIFF1 for alignment of chip to layout, than the emission image was used as TIFF2 to achieve the final localization result. In this case the various data files are shown and steps for alignment are demonstrated. Furthermore the subsequent physical failure analysis is shown.

Fig. 5 subsequently shows the layout in part a) and the two images, namely the optical reflection image b) and the emission image c.), used for the fault localization

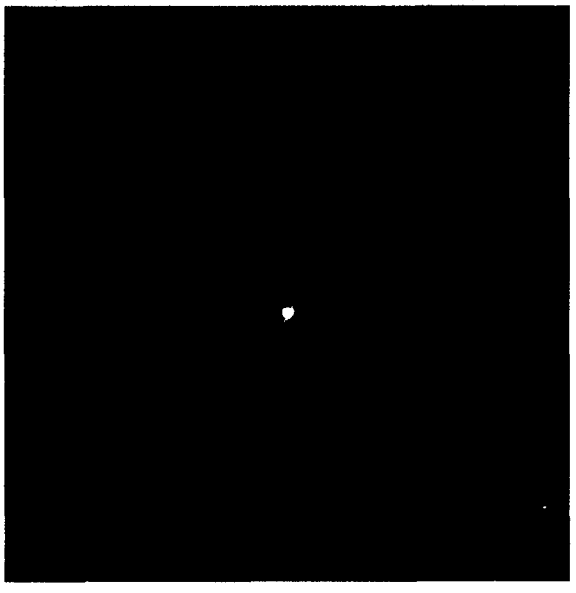


**Fig. 5c**: "Hot-spot" detected by emission microscopy

The first step of aligning the optical image inside the layout using three points of alignment is shown in fig 6. Using the emission image correlated to the layout the defect was easily localized within two poly-silicon lines of the internal ROM inside the micro-controller (see fig. 7)

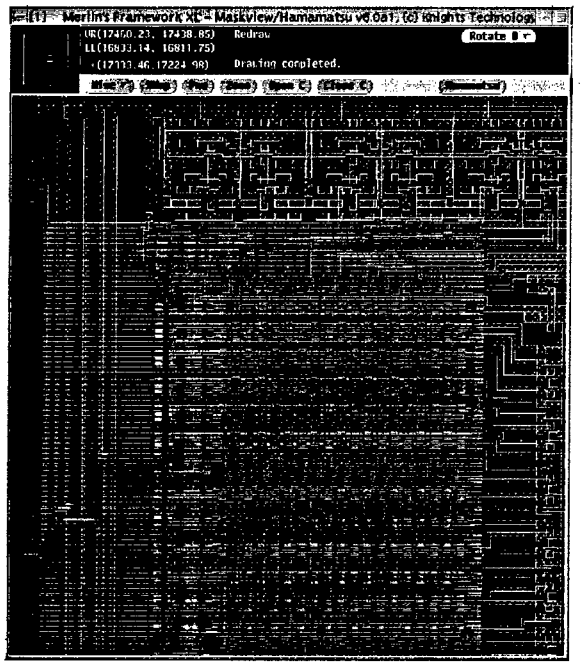



**Fig. 6**: Alignment of optical image in the layout

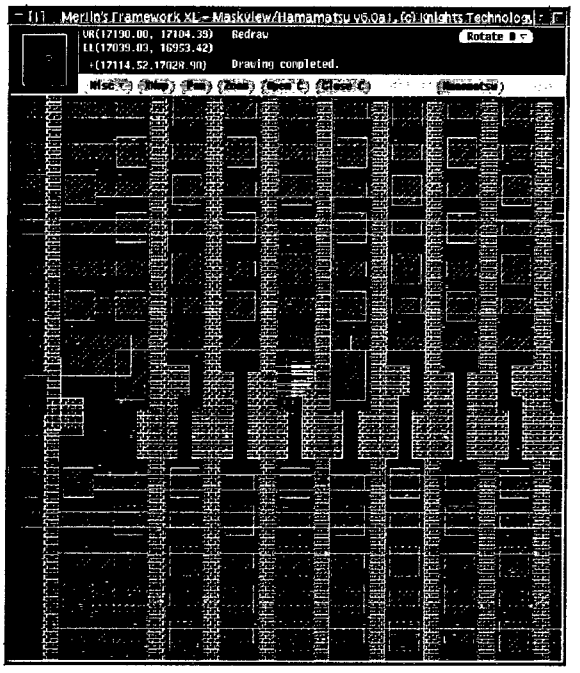



**Fig. 7**: Zoomed layout with emission spot between conduction lines of ROM

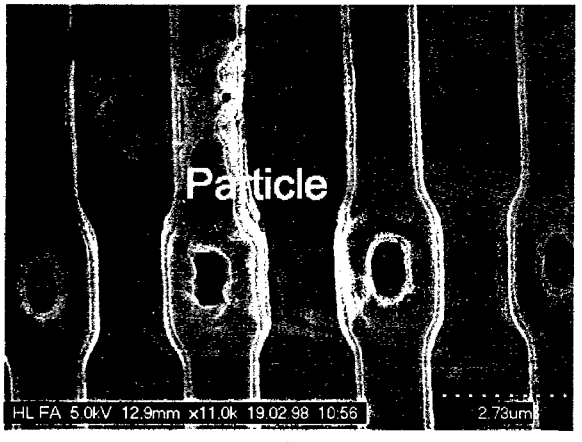



**Fig. 8**: Particle inside ROM as seen in SEM

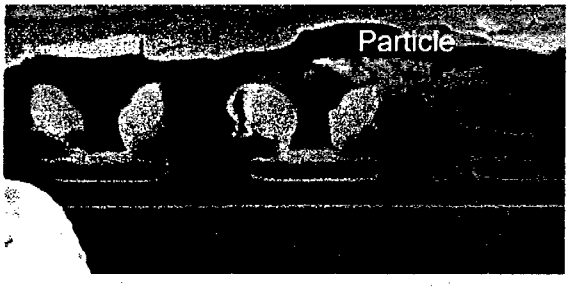


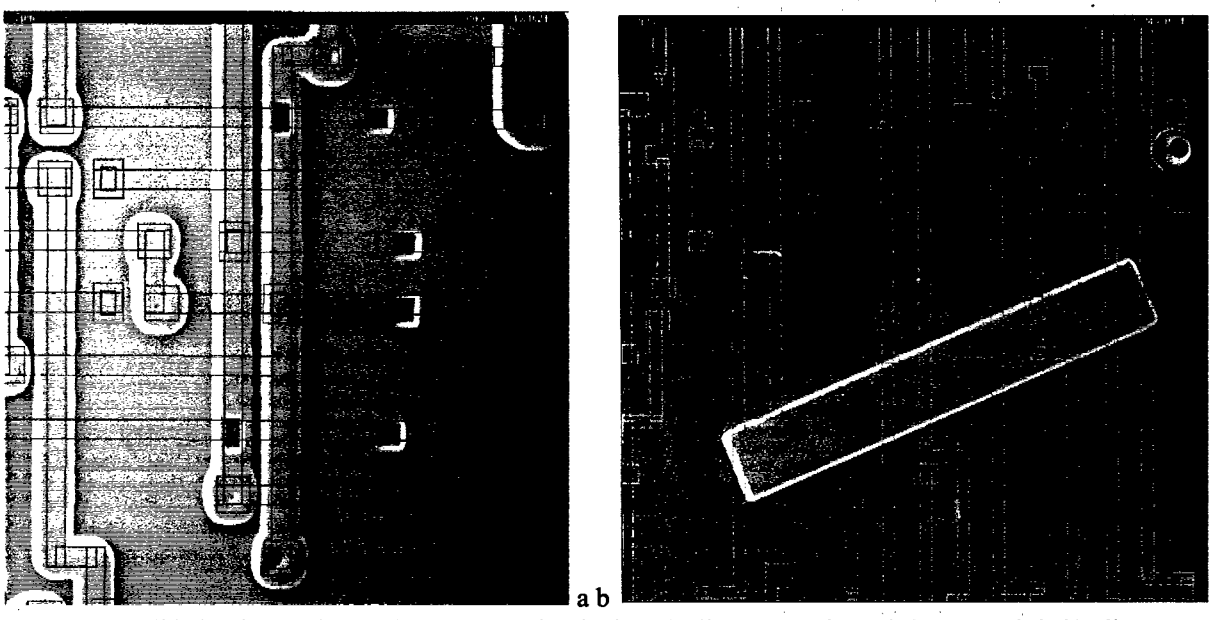

**Fig. 9**: Particle causing short between two ROM-lines as seen in FIB cross-section

After this the chip was investigated in SEM, showing a particle in the area of interest (see fig. 8). Finally in fig. 9 the cross-section explains the failure to be a short between two lines in the ROM caused by a metal particle.

In this example both the procedure and the benefits of this overlay technique are clearly demonstrated.

### *3.3 Circuit Modification Using FIB*

In the case of FIB the image overlay even helps to overcome the problem of multi wiring. In Fig 10 a) the layout was aligned using the upper layer Alu3. The layout-overlay was used to realize vias to deeper layers. Even at completely flat surfaces, showing no topography from the upper metal layers, a chip modification may succeed. At some place given some topography information (e.g. pads) the alignment is done, and than circuit modification is done using the layout overlay (as shown in fig 10 b) for an IC using CMP-technology) substituting the missing topography. After alignment the chip is modified as formerly. In fig. 10 b) new conduction is realized after gas-enhanced drilling and metal deposition (compare also Fig. 1b).



a b

**Fig. 10:** Application in FIB in modern CMP-technologies: a) alignment using Alu3 to cut Alu2, b) alignment done outside field of view, controlled by opening to Alu2

## 4. Conclusion

This layout overlay technique helps to overcome the drawbacks of modern microelectronics in failure analyses, namely navigation using optical images in sub-µm technology and navigation at surfaces having no topography due to CMP-technology. This was demonstrated in several examples. Thereby a technique was realized, which is independent from the equipment, because it uses standard image format "TIFF" and it is realized as extension of a standard CAD-navigation software.

However, this approach (like other improvements of equipment) will not solve the problems. Design for Analysis (DFA) has to be accepted as a must, as Design for Testability (DFT) is already realized. Some attempt for DFA are demonstrated [8], but more have to be realized to guarantee analyzability of modern microelectronics.

## Acknowledgement

We like to thank C. Boit and Mrs. F. Bruns for their general support. Special thanks are due to S. Shen and M. Brügel from Knights Technology for supporting all phases of this project, from definition till installation in our FA-lab. Examples are taken from our colleagues, Mrs. C. Engel and M. Matyschiok. We thank them for their cooperation.

## References

[1] St. Cocina, et al.:
"Software Integration in a Workstation-Based E-Beam Tester", Proc. ITC (1986) 644 - 649

[2] FEI Company
"Focussed Ion Beam xP Workstation",
User's Guide, Chapter 6 (1996)

[3] Micrion Corporation :
"Micrion-Knights Software Reference Manual", (7/21/97)

[4] "Merlin's Framework Users' Manual"
Knights Technology, Inc. (1997)

[5] "Equipment Independent Image Overlay"
Application Note, Knights Technology, Inc. (1997)

[6] S. S. Lee, J. S. Seo, N.S. Cho, S. Daniel:
"Application of Backside Photo and Thermal Emission Microscopy Techniques to Advanced memory Devices", Proc. ISTFA (1997) 63 - 66

[7] K. Naith, T. Ishii, J.I. Mitsuhashi:
"Investigation of Multilevel Metallization ULSI by Light Emission from Back-Side and Front-Side of the Chip", Proc. ISTFA (1997) 145 - 151

[8] N Kuji, T.Takeda:
"Automated Placement of Testing Pads for Electron-Beam Observation",
Proc. ESREF (1997), Microelectron Reliab.37 (1997) 1565-1568

PERGAMON

Microelectronics Reliability 38 (1998) 993–996

MICROELECTRONICS RELIABILITY

# Failure Analysis of Wafer using Backside OBIC Method

Seigo Ito(1) and Hideo Monma(2)

*(1)System LSI Design Dept., FUJITSU VLSI LIMITED*
*(2)LSI Manufacturing Group, FUJITSU LIMITED*
*1500, Mizono, Tado-machi, Kuwana-gun,MIE PREF. JAPAN*

**Abstract**

As a result of performing the Iddq (IC's power supply current) failure analysis on CMOS logic LSI using the backside OBIC (optical beam induced current) method, we succeeded in detecting a short circuit at an overlapped portion between a power supply line and a signal line. So far a circuit failure beneath the Al wiring has been difficult to find because the surface of LSI chip is covered by the wiring. However, we succeeded in detecting the carriers abnormally generated at a failure portion of a circuit on the chip surface by the optical energy when the backside of wafer was irradiated by an He-Ne laser. The backside OBIC technique is an effective method of failure analysis for advanced multilayer LSIs.
© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

The backside OBIC method is a technique to detect the carriers (electron-hole pair) generated by laser energy as a current flowing outside the test circuit when laser light penetrates to the silicon substrate. Until now, a failure portion has been detected by irradiating the front surface of sample chip by a laser, using the OBIC analyzer. [1,2] However, as the high density and multilayer Al wiring technology of LSI chip advance, a failure portion under the wiring becomes difficult to analyze from the front surface of the chip. [3]

Therefore, we proved experimentally that failure analysis can be done by detecting the current variation caused by the carrier generation, when the back side of a wafer is irradiated by an He-Ne 1152

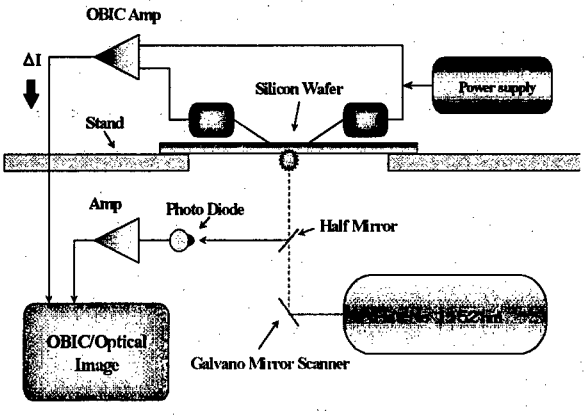



Fig. 1. Outline diagram of the backside OBIC imaging system

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00089-4

nm laser (hereafter referred to as IR laser). In this paper, we describe the effectiveness of the backside OBIC method and the results of failure analysis using this method.

## 2. Theory and backside OBIC system

To observe the front surface circuit pattern from the back side of a silicon wafer, a low energy laser which penetrates the silicon substrate is necessary. For example, considering the band-gap of silicon substrate of 1.107 eV (at 273 K), the He-Ne 1152 nm laser (1.076 eV) penetrates the silicon substrate, because no transition between the valence band and the conduction band occurs.[4] That is, the fact that the laser radiation from the back side of chip penetrates the silicon substrate, enables to view the surface circuit pattern from the back side of wafer.

If impurities are doped into the silicon substrate, the transition energy necessary to the band-gap of substrate becomes smaller than that of the intrinsic semiconductor. For example, the transition energy necessary when boron is doped into the silicon substrate decreases from 1.107 eV to 1.063 eV.

That is, with the IR laser used this time, it becomes difficult to penetrate the region where impurities were doped. On the other hand, it is easy to generate the carriers in the depletion layer. A technique to detect these carriers outside of a test circuit as a current variation is the OBIC method. The wavelength of the IR laser used this time is the optimum for penetrating the silicon substrate and generating the carriers in the surface circuit region.

The outline diagram of the backside OBIC system (JEOL JBS-1181) is shown in Fig. 1. The sample wafer was fixed at four peripheral portions by a vacuum chuck. The OBIC analysis image was created by converting the OBIC current into graphic signal. The OBIC current data was obtained by 2 dimensional scanning of IR laser irradiated the wafer backside. Out put power of IR laser is 2 mW, and scanning is made by a galvano mirror. The number of pixels was 512 x 512 pixels and it took one second to scan each graphic screen. The optical image of the circuit pattern viewed from the wafer front surface was obtained in the same manner as mentioned above, by laser scanning, receiving its reflected laser by an optical diode, and converted into a graphical image. The failure portion can be identified by overlaying the optical image with the OBIC image.

The voltage applied to the sample is supplied from the upper side of the wafer by probing. The detector sensitivity of the OBIC amplifier is 40 pA and the maximum sink current is 20 mA.

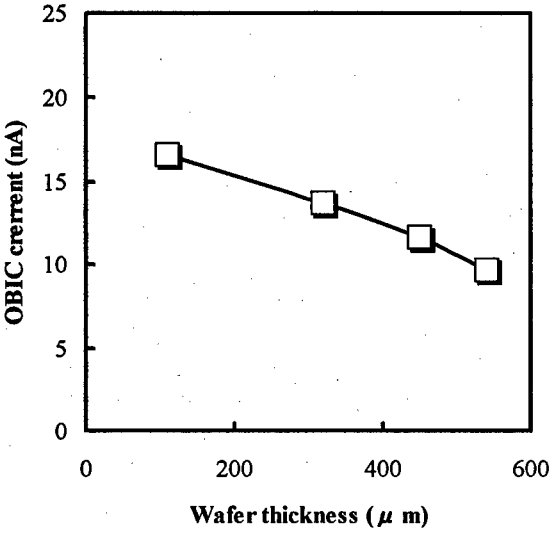



Fig. 2. Characteristics of OBIC current and wafer thickness. In the measurement, by applying reverse bias current to the p-n junction region of 10 μ m × 30 μ m, we measured the OBIC current flowing when an He-Ne laser of 1152 nm was irradiated from the back side of wafer.

## 3. Result of failure analysis

To perform the backside OBIC analysis, no special surface grinding for the back surface of wafer sample is necessary. This is because the depth of focus becomes shallow as the magnification of lens becomes higher and irregularity on the back surface of wafer can be ignored.

Fig. 2 is the graph showing the relationship between the OBIC current of the p-n junction and wafer thickness when the laser is irradiated from the back side of the wafer. Though the absolute value of the OBIC current decreases as the thickness of wafer becomes larger, the analytical image can be improved by increasing the gain of the OBIC amplifier. Considering the roughness of the back surface of the silicon wafer after a BG (back grind) process, the OBIC analysis can be performed up to the thickness of 500 $\mu$m. No special processing for wafer sample is required to perform the backside OBIC analysis.

A manufactured logic LSI's Iddq failure sample was analyzed by the backside OBIC method, using the CMOS 0.35 $\mu$m Al 3 layer wiring process.

The sample was analyzed by applying 3.3 V between a power supply line and GND using an LSI tester. It was found that a current of approximately 30 $\mu$A was flowing, which was an abnormal value compared with the normal current value of 1 $\mu$A. Considering that the circuit operation of the wafer sample was operating normally, there was a possibility of a high impedance short circuit between the power supply and the GND.

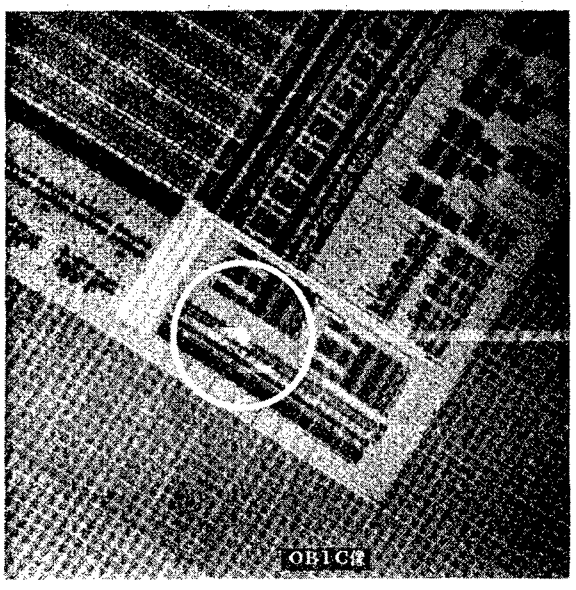



Fig. 3. Backside OBIC image 1800 magnification

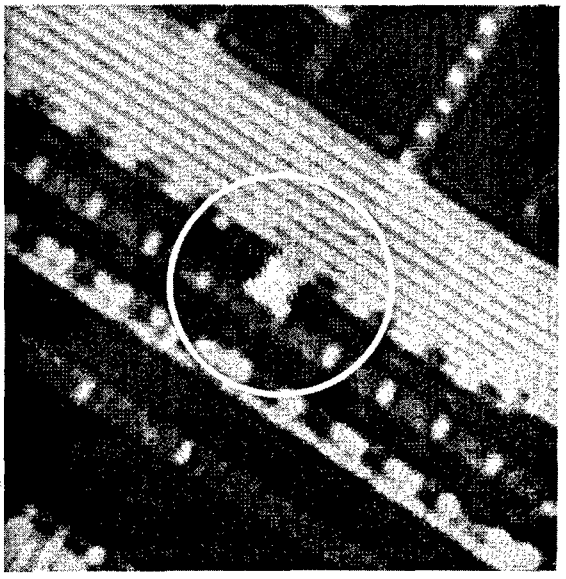


Fig. 4. Backside OBIC image 3600 magnification

The sample wafer was analyzed using the backside OBIC system and a bright spot likely to be a weak point was found in the random circuit. The OBIC current variation of the failure portion was approximately 1 $\mu$A. A bright spot was observed when a voltage of 0.1V to 0.3 V was applied between the power supply line and the GND. These results are shown in Fig. 3 & 4.

As a result of the detailed SEM investigation of the wafer sample, we found that the bright spot was not abnormal and there was parasitic resistance of approximately 100 K$\Omega$ which shunts between the inverter gate output and the power supply line. Fig. 5 shows SEM image in the Al wiring circuit short.

A strong OBIC current flows due to the reason that the bright spot of the analytical image has a higher electric field of depletion layer than other normal regions, which so far has been considered as a failure. Fig. 6 shows the equivalent circuit and cross-sectional diagram.

As the voltage for measurement used in the backside OBIC method, a fine analytical image could be obtained when the voltage is low. This is because the abnormal portion can be detected emphatically by suppressing the carrier generation at the normal portion.

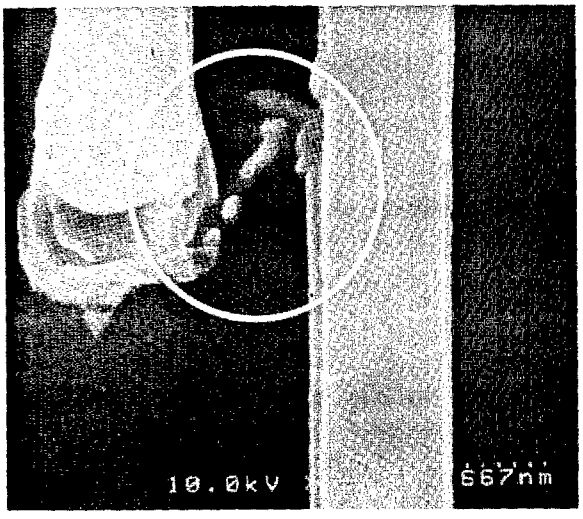



Fig. 5. SEM image of Al wiring short.

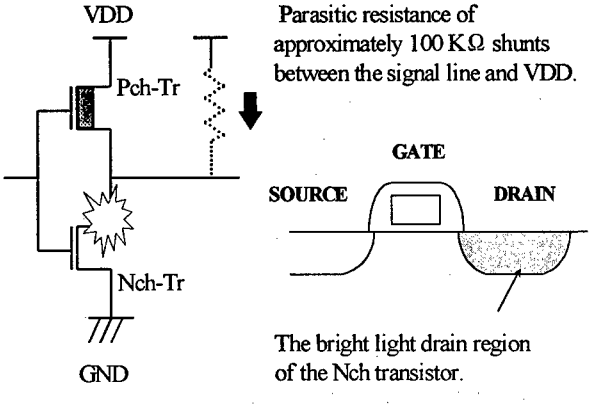



Fig. 6. CMOS circuit of the breakdown region, and cross section of the bright spot.

## 4. Conclusion

Until now, failure analysis of multilayer wiring LSIs has been difficult because the laser was intercepted by the surface Al wiring. However, we succeeded in detecting the photo-current of less than 1 $\mu$ A caused by carriers generated at a failure portion of the chip's surface circuit, by irradiating the back side of wafer sample with an He-Ne laser, using the backside OBIC method.

By applying a low voltage of 0.1 V to 0.3 V to the wafer sample, the carrier generation at normal portion is suppressed and the carrier generation (1 $\mu$ A equivalent) at the failure portion can be detected emphatically.

The silicon wafer used for the backside OBIC analysis can be analyzed despite its back surface roughness after BG process. Therefore, we don't have to prepare particularly. This is because the depth of lens focus becomes shallow and the ignored.

The wavelength of the He-Ne laser used this time is 1152 nm, which can penetrate the silicon substrate and generate the carriers in the surface circuit region. Therefore, we consider this wavelength as the optimum for the backside OBIC analysis.

## Reference

[1] T. Shiragasawa et al. 64Kbit full CMOS RAM using a laser scanner. IRPS, pp.63-68,1984.

[2] K. Haraguchi Microscopic OBIC Measurements and their Applications. IMTC, pp.639-699,1994.

[3] T. Ishii et al. Failure Analysis of ULSI with Photo emission from Backside of the chip. LSI Testing Symposium in Japan pp.221-226,1995.

[4] K. Ueda Backside OBIC Scanner. LSI Testing Symposium in Japan. pp.164-169,1996.

PERGAMON

Microelectronics Reliability 38 (1998) 997–1007

MICROELECTRONICS RELIABILITY

# ESD protection methodology for deep-sub-micron CMOS

**K. Bock, G. Groeseneken, H. E. Maes**

*IMEC , Kapeldreef 75, B-3001 Leuven, Belgium*

**Abstract**

Electrostatic discharge is considered to be a serious treat of integrated CMOS circuits since the feature size reached about 1.5-1.0μm. Since then the scaling of CMOS technologies led to an increase of their ESD susceptibility based on geometrical, physical and technological limitations. The paper describes the change in methodology in order to assure a reasonably high target value of ESD protection with newly to be developed deep sub-micron feature size technologies. The backward adaptive conservative methodology is step by step replaced by a methodology considering the ESD issue already during process development and involving more predictive ESD-TCAD into the development cycle. It is concluded that the scaling based limitations might grow to a significant problem in the near future requiring significant effort to assure a reasonable ESD protection level for CMOS technologies, in particular if the high-frequency properties of such technologies should not be affected. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Electrostatic discharge started to be considered a serious reliability hazard in CMOS technologies when feature size reached about 1.5-1.0μm. Its importance has emerged with the rush in decreasing feature size. Different kinds of electrostatic discharges can happen and there are several ESD models known to simulate reality. The most important are the Human Body Model (HBM), the Machine Model (MM) and the Charged Device Model (CDM). Commercial testers are available to test the devices and circuits for their ESD sensitivity according to these models. [1-4]

Table 1: Major NMOS Degradations caused by ESD

- •Contact spiking
- •Oxide (dielectric) breakdown
- •Filamentation
- •Thin-film burn-out
- •Metal migration
- •Interface interdiffusion

Most of the ESD related failure mechanisms (see Table 1) and their locations in a device, as depicted in Figure 1 are related to excess heat. This paper does not aim, however to discuss the various ESD models and failure mechanisms [1,2] in particular.

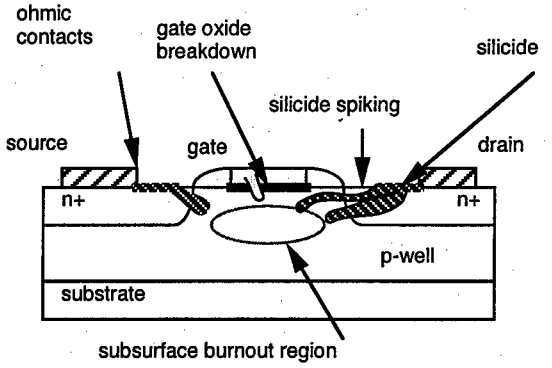



Figure 1: Locations of major NMOS degradations caused by ESD stress

The scope rather focuses on the scaling problem [5] and its impact on on-chip ESD protection and on the corresponding methodology changes in order to reach an optimised ESD protection design for a newly developed CMOS process technology. The paper discusses the measures that can be applied and the corresponding advantages and limitations. In earlier technologies the design window of ESD protection was still wide enough and a backward adaptive iteration process has been used to design a suitable ESD protection concept for a certain technology after the process development and integration phase was finished.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00148-6

Obviously such a methodology is too time consuming and can not cope with the need of reduced cycle times of today's process development pace.

At present the feature size of CMOS technologies has reached the deep-sub-micron range. As a result technologies have become highly vulnerable to excess energies introduced into the circuitry because their energy tolerance rapidly shrunk. Physical limits shrink the ESD design margin in such a way that the question can be asked how ESD protection techniques can evolve compared with the trends contended in the SIA road map [5]. As a result the CMOS ESD issue will probably emerge to be the largest reliability concern of ULSI right after or equally important to dielectrics and interconnects [6]. Moreover, dielectrics and interconnects are in the focus of ESD research as well [6]. Since the actual pace of technology scaling is even faster than predicted by Moore's law and because these near future ESD problems are not considered yet in the SIA roadmap it becomes even more urgent to have ESD solutions available, even before the process is fully developed and stable. Concluding that the possible process and technological improvements are by far not satisfactory, in a further part on-chip protection possibilities for deep-sub-micron high-frequency devices and circuits and alternative ESD protection (clamping) devices are proposed.

## 2. ESD related impact of scaling.

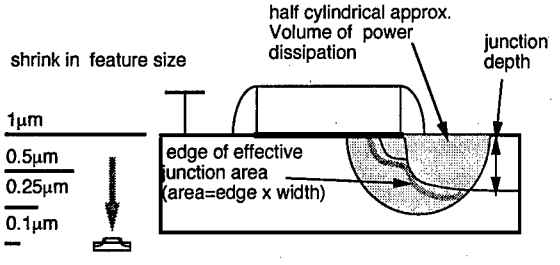



Figure 2: Illustration of the scaling of CMOS and the geometry's involved during an ESD event.

Down to feature sizes of about 1.5-1.0µm the technology development did not consider ESD issues as a major problem and correspondingly the technologies could be optimised for device properties only. Since the 2µm feature size technology however, ESD became an issue for four main reasons. 1) First of all, almost each new technology added new aspects to the set of relations between device properties and reliability concerns.

At 0.7 µm featuresize the hot carrier problem reached its top and LDD was introduced with a significant negative impact on ESD properties. At about 0.5µm feature size the device and interconnect properties, especially series resistance, forced the introduction of salicides, in particular for digital applications. This had again a very serious negative impact on ESD properties. Some new processing options, however, like the introduction of W-plugs and TiW barrier metals as an advancement in process turn out to be also beneficial for ESD.

2) Due to the scaling, the volume in which the power has to be dissipated shrunk considerably. Figure 2 illustrates the part of a NMOS protection device involved during an ESD event and the scaling from 1µm down to 0.1µm featuresize.

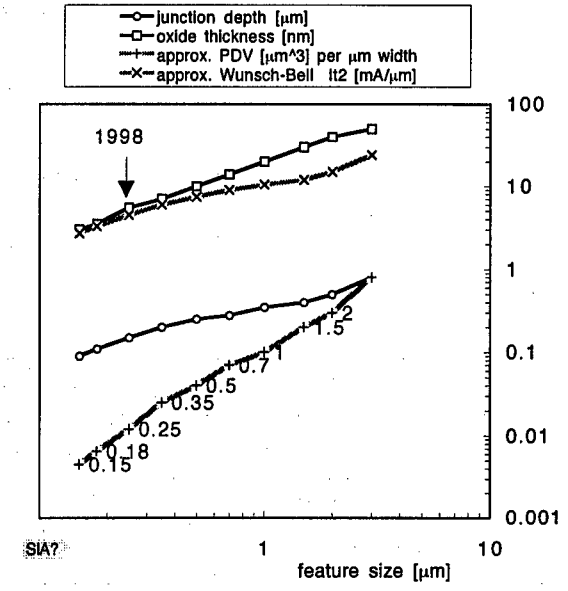



Figure 3: Scaling of junction depth, opt. oxide thickness, approx. volume of power dissipation ($PDV \sim L^3$), approx. maximum $It_2$ ($It_2 \sim L^2$) versus feature size.

Figure 3 depicts the junction depth, the optical oxide thickness, the approximated volume of power dissipation at the NMOS drain junction and the maximum physical value of the second breakdown current, $It_2$, based on a Wunsch-Bell [7,8] approximation (see Figure 4) for a NMOS structure (Figure 2) as a function of the technology feature size. The approximated effective volume of power dissipation and the approximated effective junction area during a pulse event are monotonically decreasing with feature size.

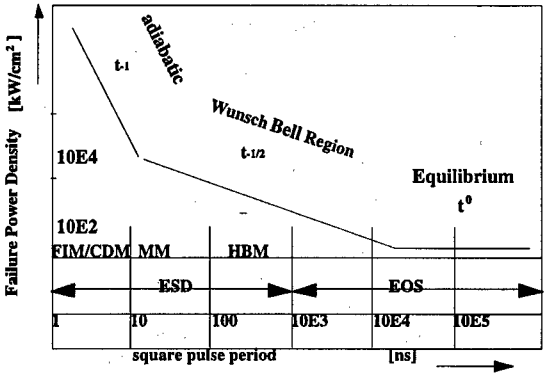



Figure 4: Illustration of Wunsch Bell failure power density (the area above the curve) versus time constant of a stress square pulse

Based on the Wunsch-Bell relation it is possible to approximately calculate the maximum power density and temperature in such a junction to occur. If the Wunsch-Bell relation and the maximum power density values for Si for a chosen pulse length of 200ns are applied on the approximated CMOS junction area of a NMOS protection device a physical maximum $It_2/\mu m$ can be derived for each technology generation[1].

The relation between power density and $It_2$ is based on the decreasing trend in holding voltage. For a possible CMOS SCR (see later) protection device this means that the reachable $It_2$ value more than doubles because of the lower holding voltage. However, these factors have a direct impact to the maximum ESD failure threshold reachable in a certain protection concept. It is obvious that this trend is decreasing with feature size and that it is not constant at all. Many known target values for NMOS based protection published so far are confirming this physical limit curve.

At feature sizes below 1μm down to 0.5μm the ESD robustness of salicided LDD CMOS processes could be maintained at reasonable values of about 5-7mA/μm $It_2$ which is comparable to about 7-10V/μm HBM [4,9-12]. The second breakdown current $It_2$ of grounded gate NMOS structures in a certain technology usually increases with a reduction of the effective channel length since this corresponds with a reduced base width of the inherent lateral bipolar transistor. It was possible to compensate the effect of shallower junction with the shorter gate length in these technologies, but only because the fundamental physical limit has not been reached yet. The fact that the $It_2$ value could be kept constant over several technologies implies that the processes have not been extensively optimised for ESD. The on-set in improvement of processes in order to reach a certain $It_2$ per μm width target value just took place at these feature sizes and therefore allowed this to happen.

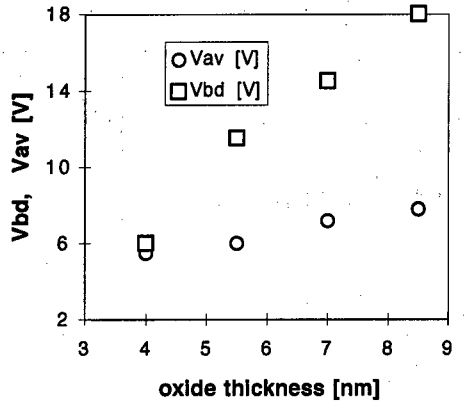



Figure 5: Junction avalanche voltage $V_{av}$ and oxide breakdown voltage $V_{bd}$ for a different oxide thickness (approx. data taken from [4]).

[1] Of course, the approximation of the junction area is influenced by tolerances in interpretation and fitting and the absolute values may be slightly higher.

However, it is clear that it is not possible to reach a constant $It_2$ target value for each new feature size technology. Very soon these constant trends in ESD $It_2$ target values will appear to be an illusion since ESD process optimisation will reach the physical limits. At this point expected to happen already at the 0.15μm CMOS generation, commonly used solid state ESD protection based on NMOS devices with tolerable area consumption will not be able to protect anymore up to reasonably high thresholds. This could grow to a problem endangering the value of the SIA road map [5].

3) A third concern is the relation between oxide breakdown and junction breakdown values [13]. Junction and oxide breakdown voltages are decreasing monotonically with feature size (see Figure 5)[4,13]. It is important for an ESD protection circuit design that the junction breakdown voltage is significantly lower than the oxide breakdown voltage. If the voltage difference would shrink to zero the use of CMOS devices for ESD protection would become impossible. It is obvious that the design window gets narrower with down scaling and that for sub-0.25μm generations this effect is possibly developing into a serious concern as depicted in Figure 5.

A relaxation of the problem is the fact that the oxide breakdown voltage is dependent on the pulse length and that the transient oxide breakdown voltage in the ESD relevant time domain (10-200ns) is significantly higher than the DC breakdown voltage. Figure 6 shows the time to breakdown TDB of an 7.1nm oxide thickness depending on the applied voltage from quasi-DC down to 10ns.

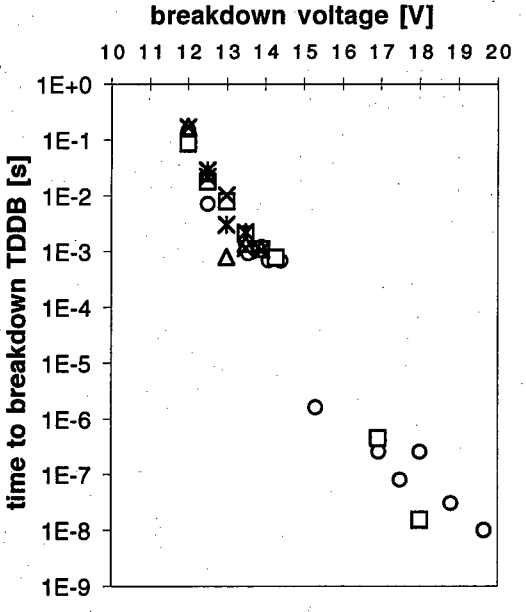



Figure 6: Time to dielectric breakdown TDDB versus breakdown voltage for a 7.1nm oxide. A more extended investigation is in progress and intended to be published elsewhere.

4) The metallisation geometry such as pitch and minimum width in a CMOS technology have a significant impact on ESD considerations. It is shown recently [14] that the ESD robustness is reduced with wire width and that a first metal M1 wire width of 2μm already reduces the HBM robustness to 4kV and that for metal level M2-M4 the width has to be 4μm to reach the same 4kV. During the last generations, ESD designers have focused on scaling of MOSFET and dielectrics, but, in the next generations interconnects will become a limiting factor in ESD design [14].

## 3. Process design options for ESD-protection

In order to build in reliability in sub-micron CMOS different options have been followed. In all cases a serious analysis of the possible process splits has to be performed. ESD contour plots for $It_2$ and snapback voltage $V_s$ and holding voltage $V_h$ should be used for the analysis of the optimisation of processing splits. If they are combined with similar contours used for the device properties they are a powerful tool to determine the trade off between ESD e.g. hot carriers and drive current. For a 0.35μm CMOS NLDD or double diffused drain DDD, ESD implants and a salicide block have been used for both a standard minimum physical gate-length performance NLDD NMOS and a 0.45 μm minimum physical gate length DDD I/O NMOS with improved ESD robustness [9]. DDD drain style provides a wider process window with space to meet performance, HCI and EOS/ESD specs. Phosphorus DDD implantation is ran after side-wall spacer etch. However, MDD (moderately doped drain) is found to be the best overall choice in [15]. Blocking the NLDD implant is a good measure to increase the ESD properties of a protection device in a CMOS technology if mask effort is not a serious argument (see Table 2). Graded junctions lead to higher snapback and holding voltage of protection devices. Abrupt junctions lead to lower snapback voltage. Another possibility is to use so-called ESD implants which are effectively partially overwriting a graded NLDD with the help of a second implantation[1,9,15-18].

Table 2: Overview of the DC-snapback parameters for 50μm wide NMOS reference devices R1 - R6, SWS=0 means distance source to well contact, FS means fully silicided, SB means silicide blocked, NLDD means NLDD implanted (otherwise blocked).

| | Vava [V] | Vt1 [V] | It1 [mA] | Rsub [ohms] | Vh [V] |
|---|---|---|---|---|---|
| R1, LDD, FS | 11.07 | 12.56 | 1.74 | 794 | 7.45 |
| R2, LDD | 10.90 | 12.54 | 1.45 | 990 | 7.82 |
| R3, LDD, SB | 10.89 | 12.33 | 1.20 | 1066 | 7.26 |
| R4, FS | 10.09 | 11.32 | 3.63 | 239 | 6.19 |
| R5 | 9.86 | 11.24 | 3.63 | 223 | 6.38 |
| R6, SB | 9.75 | 11.14 | 2.28 | 313 | 6.10 |

This can be before or after spacer deposition and allows adjustment of the ESD related properties at the cost of device performance. For a fixed drain architecture typically the application of epi-layers results in an increase of the snapback and holding voltage. The thinner the epi-layer the higher the voltages. This is related directly to the decrease of the emitter base junction shunt resistor of the inherent lateral bipolar transistor.

Lower ESD degradation thresholds are obtained for fully salicided CMOS processes, because they lack the inner ballast resistance. If the salicide thickness is increased the ESD threshold dramatically decreases. Therefore, for a consistently good ESD protection level, there is a maximum limit on the salicide thickness formed on a shallow junction [16]. The thickness is reported to be less than that required to ensure an optimum compromise between a low junction leakage current, which also needs a thin salicide layer and a low series resistance asking for thicker salicidation. This is leading again to a trade-off between device performance and ESD tolerance [16,17]. It has been shown that if the junction depth is made deep enough by an extra implantation to compensate for the effect of the salicide the drawback in ESD performance can be prevented to at the price of device performance and additional process steps. A possibility of preventing the negative salicide effect at least to the input and ESD protection transistors is to use a salicide blocking at the price of additional mask and process steps.

The influence of drain engineering, in particular the positive influence of higher arsenic drain/source implantation energies towards good device second breakdown current values for a fully silicided 0.25μm CMOS technology has been shown recently [19]. The influence of a retrograde well profile on the ESD performance is used to improve the built-in reliability of a 0.25μm CMOS process technology [19].

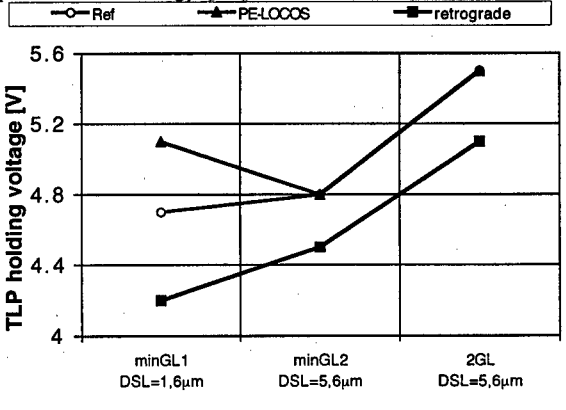



Figure 7: Holding voltage of fully-silicided NMOS devices with three different DSL (drain silicide length) and two different gate length for process splits comparing a flat well with two different isolation schemes (ref (PBL), PELOCOS) and a retrograde well [19].

The ESD relevant parameters such as the snapback trigger ($V_{t1}$) and holding voltage ($V_h$) (see Figure 7) and the snapback trigger ($I_{t1}$) and second breakdown current ($I_{t2}$) as well as the HBM thresholds are found to be positively influenced by the retrograde well doping profile.

## 4. Device design options for ESD-protection

Even if all these process tunings are conside red the ESD failure threshold especially of fully silicided technologies can be very small and additional ESD protection clamps and circuits (see Figure 8) have to be foreseen in order to protect the circuitry from the hazardous ESD [20-27]. Electrical on-chip ESD protection can be based on different types of clamping elements like Diodes, Zeners [26], avalanche diodes, NMOS based snap-back structures (NPN) (see Figure 9), silicon controlled rectifier "SCR"[24] (see Figure 10).

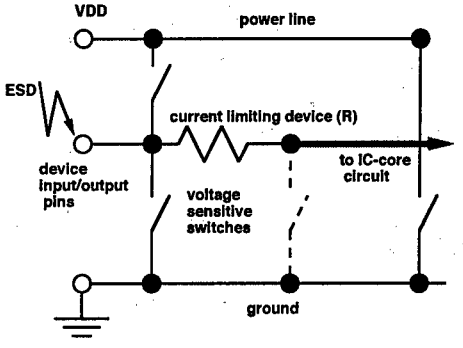



Figure 8: Basic ESD protection schematics.

For the deep sub-micron in particular, and for low-voltage circuits [23,25], diodes are suitable as protection devices because they can be used as a stack in forward polarity [27]. Forward biased diodes are minority carrier devices after turn-on. In reverse polarity the power dissipation of diodes is too high because they are majority carrier devices beyond avalanche, and the current flow is mainly determined by the field. As a result they have very poor protection property in reverse polarity.

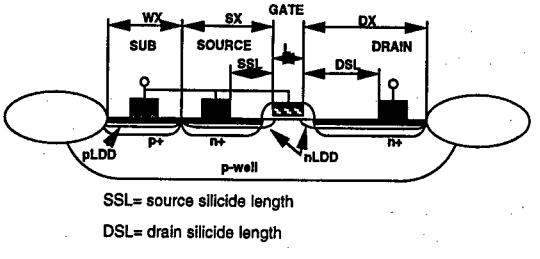



Figure 9: NMOS in grounded-gate configuration with labelling of device geometry's.

Zeners are often used as triggering or coupling devices in dynamically gate-controlled configurations[26]. They are based on hard $N^+/P^+$ junctions and need a special process option to be available in CMOS processes.

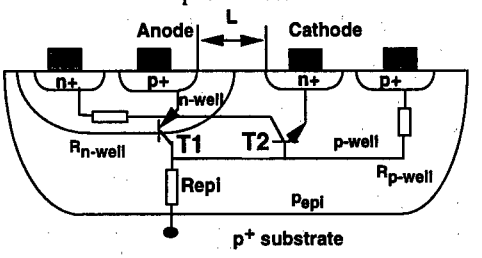



Figure 10: SCR schematics in a CMOS technology

Thick oxide or field oxide devices are very ineffective in the sub-micron range whereas they have had very good properties in the micron feature size range, better than thin-oxide devices. Thin oxide devices are now performing better in the sub-micron regime. Caused by the shorter gate length the inherent bipolar becomes more effective leading to an increase in failure threshold by geometrical advantage only in thin oxide devices [1]. Gate-coupled thin-oxide structures are performing well if the trigger inhomogeneities are solved [20,21]. A possibility to force homogeneous triggering of NMOS structures is provided by a dynamically controlled gate or gate-coupling. The aim is to design a coupling of the ESD pulse (time-constant $\tau$ = RC of about 15-50 ns) over a coupling device directly to the gate of the NMOS in order to enhance the uniformity and the speed of triggering. The capacitance couples for a short time a positive voltage to the gate. This short time MOS mode causes the establishment of a homogeneous carrier distribution in the MOS. The well adjusted RC constant allows an increased failure threshold to be reached mainly based on a more uniform triggering of the device. If the gate voltage is then pulled to ground by discharge of the capacitance through the resistive part these carriers contribute to the current through the NPN. If the coupling pulse is well adjusted there is no avalanche and snapback needed to trigger the bipolar transistor (reduced snapback voltage) and a more homogeneous current flow is established in the NPN [4,20,21].

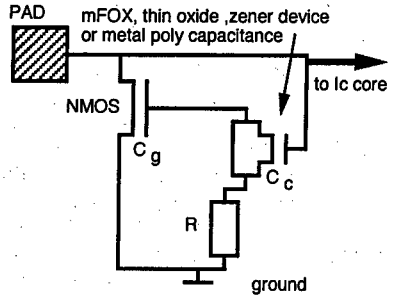



Figure 11: Schematics for gate coupling of an NMOS ESD protection clamp.

Figure 11 shows a basic coupling circuitry with the capacitance Cc of the coupling device and a resistors R. R ensures that the capacitance discharges

if the coupling device itself stays too high ohmic during the pulse. The area of the coupling element and the value of R are used to tune the time constant. The coupling capacitance $C_C$ should not be much smaller than the capacitance $C_G$ of the device to be coupled.

SCR's (silicon controlled rectifier, see figure 9)[1,4,24] - a kind of thyristor - could lose its importance in the very deep sub-micron technologies. The holding voltage is determined by the p-substrate resistance. The anode to cathode clamping voltage determining the SCR holding voltage can be as low as 1-2 volts (two forward biased diodes) leading to very low power dissipation in the SCR clamp. The trigger voltage of the SCR is controlled by the n-well overlap of the $p^+$ anode and is normally very high. In the deep-sub-micron technologies now the SCR action condition $ß_{npn}ß_{pnp} \geq 1$ is influenced by technological changes. ß is smaller for shallower junctions causing SCR's to become less efficient in advanced CMOS. Additionally the turn-on voltage is increased with decreased $R_{sub}$ and $R_{n\text{-well}}$. Consequently a thicker epitaxial layer would be recommended for increased ESD performance, leading to a trade -off with the latch-up problem optimised with thinner epitaxial layers. There are different techniques presented to lower the SCR trigger voltage[1]. NMOS triggered, gate coupled and Zener-triggered SCR's trigger at voltages less than 10 V. For HBM and for high-ohmic input structures they can provide a very high protection level.

For output transistors the breakdown voltage of the NMOS used as drivers is lower and therefore a resistor is needed to directly decouple the protection circuit from the output driver degrading the performance of the output circuit. Further the dynamics of the SCR is limiting its protection ability for CDM and other very fast transient stress. However, this could be solved by combining the SCR with a grounded gate or gate coupled NMOS protection clamp ,but, increasing the area and effort for the ESD protection circuit additionally.

## 5. ESD protection design methodology

The task of designing ESD protection has become a challenging objective [all]. On the one hand faster development cycles require very early actions already during the process module development and process integration phases. This also induces the more backward adaptive nature of conservative ESD protection design ( see Figure 12) to be transformed into a more predictive methodology. Moreover including ESD reliability relevant process, device and circuit simulation in parallel to the process development (see Figure 13) is needed in order to optimise the well-known trade-off between the important device properties like drive current and threshold voltage on one side and the reliability properties of a process like hot carrier life-time issues and ESD withstand thresholds on the other side. On the other hand the availability of such quantitative simulation tools is still limited.

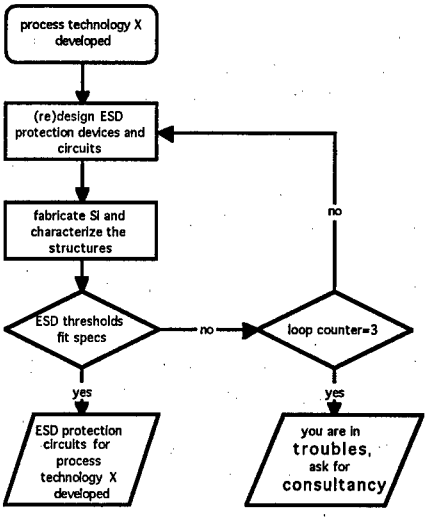



Figure 12: Conservative back-ward adaptive development of ESD protection.

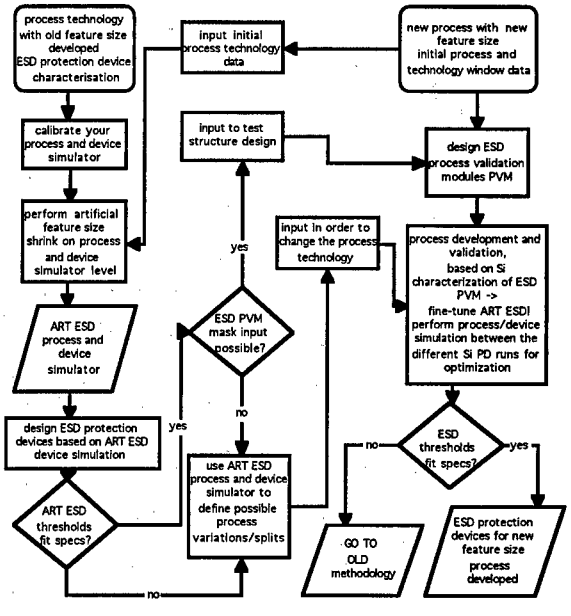



Figure 13: Advanced Methodology including CAD into a CMOS process development in order to receive a predictive approach resulting in a parallel development of process technology and ESD-protection.

### *5. 1. Classical backward adaptive design approach*

In the classical methodology, design and optimisation of ESD protection are performed by first ESD intensively testing of the devices or circuits to be protected. Based on these measurements the distributions and average values in the ESD-tests are defined. Test for possible process tolerance influence (at least two different lots)

has to be performed. Based on these test results the design of the ESD protection is performed with respect to the lower edge of the distribution and by following certain ESD guidelines as collected in Table 3 and Table 4. Selected protection elements have to be characterised in a next step . This means that all types of ESD stress have to be considered. Here it is now important to regard the fact that the width scaling of the ESD failure threshold of the transistors is limited. About 75μm for a 0.25μm CMOS fully silicided process has been reported [17].

Table 3

ESD protection techniques itemised

- **Optimise:**
  - Contact distance from the heat dissipating junctions
  - Junction area and periphery (as large as possible)
  - Clamp voltage over dielectrics and oxides
  - Add resistance (drain side) to defocus current and to keep dissipation away from the contacts
  - Ensure uniform current flow. Current flow can be influenced by the implantation condition species, dose and energy
  - Adequate via and contact design
- **Avoid:**
  - current concentration and hot spots
  - Corners
  - Narrow metal lines (! $R_{bus}$)
  - Sharp bends in (poly) resistors

Table 4

ESD protection strategy itemised

- Power dissipation off-chip
- Low current density design
- Maximise heat dissipating areas
- Low voltage across dielectrics and oxides
- Prevent heating of the contacts
- Find (kill or use) parasitic current paths
- **Strengthen components to be protected (self protection) before adding extra protection structures or circuits**

Therefore the design of protection structures for deep-sub-micron has to be based on multi-finger structures and then also the triggering of all fingers should be assured by a symmetrical design. With longer gate length, devices in a certain technology are reaching only very low $It_2$. For example output transistors and current sources are designed with longer effective gate length and are critical in deep-sub-micron technologies.

### 5. 2. *Predictive ESD protection methodology based on TCAD*

Starting point for the establishment of an ESD protection design methodology based on TCAD [28] is an accurate simulation of the process [28]. The activity proceeds in three steps. A first step is the implementation of the CMOS process flow in the simulator. A second step is the validation of the simulations using split information and data obtained from the process in a development stage. The third step is the calibration of the simulator to recent measurement data. Based on the process simulation device simulations are performed.

After the process has been calibrated, a feature size shrink can be performed on the artificial (ART) device simulator level by adapting to the initial process conditions of the new process technology to be developed. After a process simulation with a shrunk mask layout a device trend simulation can be performed with the ART device simulator. Different process variations can be simulated and corresponding device simulations are based on these process split trends. Therefore, after the process development is finalised the layout considerations of possible ESD protection clamps and circuits should be well prepared. If not, the old backward adaptive approach has to be started.

### 5. 3. *ESD Device/Circuit Simulation*

The integration of CAD software in the ESD protection design process is an area of growing concern and necessity [28,29]. From a pure technical point of view it is possible to simulate the electrothermal behaviour of microdevices during ESD events, but, this is still far from the predictive quantitative simulation of the operation of microdevices under ESD conditions. It is obvious that a predictive CAD based methodology can lead to reduced development costs and less time-consuming testing resulting in a final computer aided development of built-in reliability devices [28]. The value of device simulation under normal operation mode is undoubted but for the ESD regime several basic differences have to be considered relativating the usefulness of CAD for ESD issues at present. The consequent questions are: What is available? and Where to adapt for ESD? The basic theory is available inclusive high-field and high-current behaviour and effects. Simulation tools and CAD platforms for electrothermal simulation are available as well. Accurate device models are essential for all levels of CAD modelling. First attempts to derive macromodels (compact models) at device and circuit level for full electrothermal circuit simulation are still limited but promising. The basic nature of electrostatic discharges can be described by a very high current and very short duration injection.

In particular, physical parameter models (see Table 3) are hardly validated at elevated temperatures and for high-current injection and therefore are still unsatisfactory for ESD Simulations. In order to simulate ESD phenomena these physical parameter models have to be significantly extended in their validity range. Exact data of some electrothermal material properties [28] are difficult to access since they are process dependent and they are necessary for reliable quantitative analysis [28]. Device simulation is based on process simulation, and the accuracy of process simulation is connected to the accuracy of technology models available.

At present several technology steps can not be described in a way which would be satisfactory for ESD relevant device simulations. An example is the salicidation process [30,31] used to reduce the series resistance. Beside its device property related importance salicide has also major influence on the ESD properties. At present, the resistance of salicides cannot be predicted satisfactory in most process simulator software. The salicide process as well as the layer properties are not yet modelled in a physical way. A first approximation of the salicide influence is to consider its finite resistance. Some simulators are able to consider the salicide processes for 1-D process simulation and recently with 2-D possibility based on such a homogeneous material property assumption [28]. For a correct electrical simulation of the ESD device it is believed that the current spreading due to a finite resistance of the salicide layer plays an important role.

Table 5
Physical parameter models which need to be adapted for ESD simulations

| |
|---|
| Effective Intrinsic Concentration |
| Band Gap effects |
| Effective Density of States |
| Mobility |
| Acoustic Phonon Scattering |
| Ionised Impurity Scattering |
| Carrier-Carrier Scattering |
| Surface Scattering |
| Drift Velocity Saturation |
| Generation-Recombination rates |
| Shockley-Read-Hall model parameters |
| Auger Recombination |
| Impact Ionisation |
| Thermoelectric Power (Seebeck coefficients) |
| Melting Points of semiconductor and related materials |

A second step could be to consider the heterogeneous material properties of the salicide layer as there are different phases and orientations as well as differences in grain and grain boundary properties. These microscopic nonhomogeneities are responsible for the degradation of the interface between the salicide and the silicon. Failure analysis shows that the salicide edge is a degradation centre and 3-D simulation will be needed to study these phenomena.

A calibration of simulation models is usually done by matching to measured and simulated DC output characteristics. Since the physical models and properties at high temperature and high injection are lacking, the calibration can be done by comparison to the TLP characteristics of selected devices. The TLP characteristic "contents" this dependence of the devices to the ESD related operation range, but, in a very limited way only valid for the geometries measured. As long as this is not overcome, silicon is always needed for the calibration process and a predictive methodology of ESD protection design only based on CAD is limited to trends only.

In parallel the ART based device simulations can be used to extract macro models for circuit simulation [32-37] in order to optimise ESD protection circuitry like gate-coupled or gate-controlled structures. Device simulations give precise insight into the device's internal local variables such as the electric field or the current density, but can not easily be applied for optimising extended devices or whole protection circuits because of the time-consuming nature of finite difference or Monte-Carlo based physical device simulations.

Studies of ESD (CDM) related device phenomena or ESD trigger issues were made in [34, 35, 45-48] by both device simulations and experiments. Compact models used in a circuit simulator can give fast feedback to the circuit designer. Various compact ggNMOS models for ESD are proposed [32-37]. They allow thermal considerations [32-36] and describe snapback behaviour of the parasitic bipolar transistor [36, 38-44]. At present the models [32, 33, 37, 42] have been mainly applied to Human Body Model (HBM) ESD or Electrical Overstress (EOS) in general. There are not many publications available where compact modelling considers CDM discharges, reaching amplitudes of several ampere of current within a nanosecond time scale.

Because of the limited possibility of compact models to deal with the partially three dimensional physical problems of ESD, 'hybrid' approaches are a possible way to continue. Circuit simulation with models considering device internal geometrical effects and device simulation considering external circuitry could be suitable combinations as long as we have to agree on a compromise between simulation time and accuracy. Even if the device simulation can not be as accurate as needed because of the above described lack of the physical model accuracy a qualitative trend analysis of device/technology behaviour is very helpful in the next generations of sub-micron CMOS processes to be developed. There is a need to extract common technology influences to ESD in order to receive reference points for technology development. It would also help to ease comparison of published results and to understand the influence of geometrical scaling factors, device design and layout option as common trends in a kind of ESD sensitivity analysis. Vehicles for this kind of ESD trend research should be the major building blocks of devices and the major processing modules of ULSI CMOS.

### 5. 4. *High frequency optimisation*

A key concern of the methodology is the fact that fast increase in operation frequency requires adaptation of the ESD protection principles [49,50]. In larger micron feature size technologies ESD protection circuits based on various clamping devices have been added in parallel to the functional circuitry in order to provide an alternative path for the hazardous ESD energies in case of an ESD event. In the deep sub-micron CMOS technology very high operation frequencies can be reached and are needed to fulfil the rising request of speed and functionality.

If the total input capacitance becomes too large high-frequency signals are loaded to ground. For a typical IC input normally several protection switches (clamps) have to be applied in a complete protection circuit, which leads to a serious increase of the shunt capacitance of the HF-inputs of the circuit. An ESD protection circuit for the microwave regime [49,50] has to switch about a decade faster than the devices to be protected, they have to be low-parasitic and therefore to be able to switch high signal levels and consequently very high current densities. Just adding a common clamping circuitry would reduce the input admittance for high frequencies so much that the input signal would be loaded down[50].

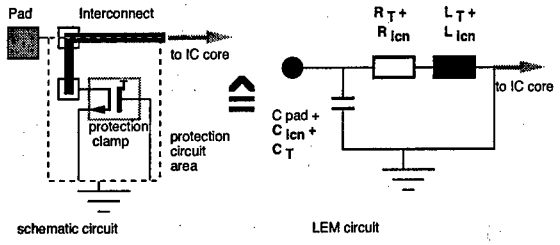



Figure 14: ESD protection is always increasing the parasitics in a circuit. As higher the operation frequency as more critical this affects the circuit functionality.

Hence, a trade-off between the implementation of satisfactory protection level and the limitation of the high-frequency device and circuit properties is entering the scene [50]. Consequently different concepts will be needed for the HF protection of deep-sub-micron technologies. One possibility is to design the output driver structure to be self-protective up to a basic ESD protection level. If this level is reached a current "booster" clamp opens in order to limit the current stress in the HF-output. Such a concept (see schematics in Figure 15) has the advantage that there is less redundant Si area consumption. The clamp does not need to trigger at a certain voltage ,but, by a current level in the output line.

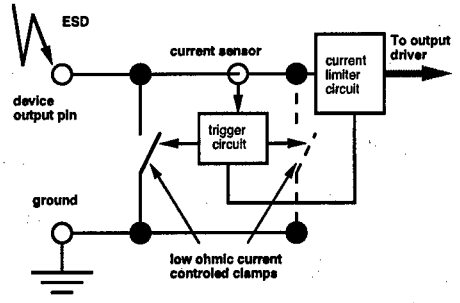



Figure 15: HF-ESD protection principle with current triggered booster clamps preventing the decoupling output path resistor (R see Figure 8) of commonly applied output ESD protection circuits. The clamp size can be reduced and the decoupling with resistors, disturbing especially the output circuits, could be prevented to a large extent. In order to optimise the HF circuits it is needed to study the building-block circuits like buffers, I/O drivers, amps, clamps and the device level building blocks like contact, junction, metal-interconnect , gate-oxide and inherent bipolar behaviour in order to optimise the weakest parts resulting in increased self-protection level of a technology. For HF circuits it is obvious that the area of the ESD protection devices and also the effect of additional interconnecting transmission lines have to be as small as possible (see figure 11) [50] and that the protection device itself has to be highly efficient in order to provide a reasonable protection level with such a small protection device. Monolithically integrated solid-state circuits, tuned to reach the highest frequencies can not at the same time be tuned to meet the highest power switching capabilities.

### 5. 5. *New protection device concepts*

Field-emission devices are proposed as a possible solution (see Figure 16) [49,50].

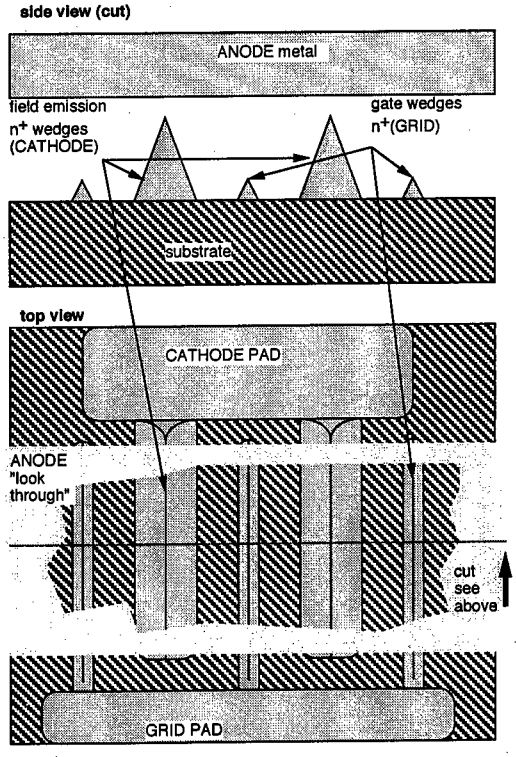



Figure 16 Schematics of a co-planar field emission triode as proposed in [50].

Field emission could be the physical principle to be applied when thinking about a non solid-state ESD protection switch based on microscopic monolithically on-chip integrated structures for ESD protection [50].The advantage of an active electrical device with a vacuum channel is very well known from the high frequency tubes. The dielectric constant $\varepsilon_r$ is 1 as compared to $\varepsilon_r$=10-13 for most semiconductors. As a consequence parasitic capacitancesof the protection elements scale down with this relation. There is no power

dissipation in the vacuum channel between cathode and anode. The limit of power handling with a field emission device (see [49,50]) is determined by the cathode and anode design. The cathode has to emit the high current density reliably. For conventional vacuum tubes, where the cathode is heated (thermo-emission), this is one of the limiting problems. For the cold cathodes (field-emission) the electrons are extracted only by applying a field strong enough to mainly reduce the work-function barrier at the cathode surface increasing the probability that electrons can tunnel through it. Therefore the problem of overheating is shifted to much higher current densities as compared to the conventional vacuum tubes. Additionally there seems to be a much higher maximum solid-state current density applicable for nm-scale structures, which is shifting the maximum value of current density able to be emitted out of a field emitter to values up to $10^8$ A/cm$^2$ [50]. Field emission elements can be built in a CMOS process and be constructed with lowest parasitic capacitance (<< 0.1pF). They reach switching times much lower than 1ps, leading to operation frequencies up to THz, since the field emission effect is instantaneous and is not delayed by carrier multiplication as compared to the gas discharge.

## 6. Summary

Electrostatic discharge is considered to be a serious treat to integrated CMOS circuits since the feature size reached about 1.5-1.0μm. Since then the scaling of CMOS technologies has led to an increase of their ESD susceptibility based on geometrical/ physical and technological limitations. Dielectrics and metallisation issues are also expected to cause potential ESD problems in the near future. Because of the very serious nature of such ESD related scaling limitations they should be considered by the SIA road map. The paper describes the change in methodology in order to assure a reasonably high target value of ESD protection with newly to be developed deep-sub-micron feature size technologies. The backward adaptive conservative methodology is step by step replaced by a methodology considering the ESD issue already during process development and involving more predictive ESD-TCAD into the development cycle. It is concluded that the scaling based limitations might grow to a significant problem in the near future, requiring serious effort to assure a reasonable ESD protection level for 0.15μm technologies and smaller in particular if the high-frequency properties of such technologies are needed. Field-emission devices and structures are a possible solution for advanced ESD protection of advanced CMOS in the deep-sub-micron.

## Acknowledgements

The authors would like to thank the members of the EU ESPRIT project ESD Design Methodologies (ESDEM) No.23643, in particular ATIS/FhG, Bosch, ETHZ, ISE, UBO and ST for numerous scientific discussions and the European Union for the funding.

Robin Degraeve for fruitful scientific discussions about oxide breakdown and for his support during transient oxide breakdown measurements, Christian Russ for fruitful scientific discussions, Vesselin Vassilev for carefully proof reading the manuscript and for fruitful scientific discussions and Bart Keppens for his excellent technical support.

## References

[1] Amerasekera A., and Duvvury C., ESD in Silicon Integrated Circuits, Wiley, New York, (1995).
[2] McAteer O., An effective ESD awareness training program, 2nd EOS/ESD Symposium (1980), pp.1-6.
[3] Duvvury C., Amerasekera A. ESD : a pervasive reliability concern for IC technologies, Proc. IEEE 81,(1993), 690-702.
[4] Duvvury C.; Amerasekera A. State-of-the-art issues for technology and circuit design of ESD protection in CMOS ICs in Semiconductor-Science-and-Technology. vol.11, no.6; June (1996); pp.833-847.
[5] SIA Semiconductor Technology, Workshop Group Reports, (1993), p.3
[6] Sematech's Reliability, TAB, ESD working group (1996)
[7] Wunsch D.C. and Bell R.R., IEEE Trans. Nuc. Sci. VOLNS-15, 6, Dec. (1968), pp.244-8
[8] Siemsen K.,1987 9th EOS/ESD Symposium, pp.168-173.
[9] Diaz C., Kopley T., Marcoux P., Building-in ESD/EOS reliability for sub-halfmicron CMOS processes, Proc. 33rd IRPS, (1995), pp.276-83.
[10] Carbajal B.G.-III, Cline R.A., Andresen B.H., A successful HBM ESD protection circuit for micron and sub-micron level CMOS,Electrical Overstress/Electrostatic Discharge Symposium Proceedings. EOS/ESD Assoc, Rome, NY, USA,(1992), 5B.2/1-9.
[11] Cook C., Daniel S. , Characterization and failure analysis of advanced CMOS sub-micron ESD protection structures, Electrical Overstress/Electrostatic Discharge Symposium Proceedings. EOS/ESD Assoc, Rome, NY, USA,(1992), 4.4/1-9.
[12] Ramaswamy S., Duvvury C., Kang S.M., EOS/ESD reliability of deep sub-micron NMOS protection devices, 1995 IEEE International Reliability Physics Proceedings. 33rd Annual (Cat. No.95CH3471-0). IEEE, New York, NY, USA,(1995), 284-91.
[13] Lin D.L., ESD sensitivity and VLSI technology trends, thermal breakdown and dielectric breakdown, Proc. 15Th EOS/ESD Symposium, (1993), pp.81-93.
[14] Voldman S., ESD Robustness and Scaling Implications of Aluminium and Copper Interconnects in Advanced Semiconductor Technology, Proc. 20th EOS/ESD Symposium,(1997), pp.316-329.
[15] Wei Y., Loh Y., Wang C., Hu C., MOSFET drain engineering for ESD performance, Electrical Overstress/Electrostatic Discharge Symposium Proceedings. EOS/ESD Assoc, Rome, NY, USA, (1992),4.3/1-6.
[16] Chen, K.L., Effects of Interconnect Process and Snapback Voltage on the ESD Failure Threshold of NMOS Transistors, Proc. 10th EOS/ESD Symposium, (1988),
[17] Amerasekera A., Correlating drain Junction Scaling, Salicide Thickness, and Lateral NPN Behavior, with the ESD/EOS Performance of a 0.25μm CMOS Process, IEDM 1996 Symposium

proceedings, p893-896, ISBN 0-7803-3393-4, 1996.
[18] Consiglio R., Huang T., ESD reliability impact of P+ pocket implant on double implanted NLDD MOSFET, Proc. 17th EOS/ESD Symposium, (1995), pp.
[19] Bock K., Russ C., Badenes G., Groeseneken G., Deferm L., Influence of Well Profile and Gate Length on the ESD Performance of a Fully Silicided 0.25µm CMOS Technology, Proc. 20th EOS/ESD Symposium,(1997), pp.308-317.
[20] Chen J., Amerasekera A.,Duvvury C., Design Methodology for Optimizing Gate Driven ESD Protection Circuits in submicron CMOS Processes, Proc. 20th EOS/ESD Symposium,(1997), pp.230-239.
[21] Voldman S., Assaderaghi F., Mandelman J., Hsu L., Shahidi G., Dynamic Threshold Body- and Gate-Coupled SOI ESD Protection Networks, Proc. 20th EOS/ESD Symposium,(1997), pp.210-220.
[22] Rountree R., Hutchins C.L., NMOS protection circuitry, IEEE Trans. Electron Devices, 32, (1985), pp.910-917.
[23] Worley E.R., Gupta R., Jones B., Kjar R., Nguyen C., Tennyson M., Sub-micron chip ESD protection schemes which avoid avalanching junctions, Electrical Overstress/Electrostatic Discharge Symposium Proceedings (IEEE Cat. No.95TH8088). ESD Assoc, Rome, NY, USA, (1995), 13-20.
[24] Sharma M., Campbell J., Choe H., Kuo C., Prinz E., Raghunathan R., Gardner P., Avery L., An ESD protection scheme for deep sub-micron ULSI circuits,1995 Symposium on VLSI Technology. Digest of Technical Papers (IEEE Cat. No.95CH35781). Japan Soc. Appl. Phys, Tokyo, Japan, (1995), 85-6.
[25] Krakauer D., ESD protection in a 3.3 V sub-micron silicided CMOS technology, Journal-of-Electrostatics.,Vol.31,no.2-3, Dec.(1993), 111-29.
[26] Richier C., Maene N., Mabboux G., Bellens R., "Comparison of Different On-Chip ESD Protection Structures in a 0.35µm CMOS Tech- nology", in Proc. ESREF Symp., (1997), pp. 1537-1540.
[27] Voldman S., Gerosa G., Gross V. P., Dickson N., Furkay S., Slinkman J., "Analysis of Snubber-Clamped Diode-String Mixed Voltage Interface ESD Protection Network for Advanced Microprocessors", in Proc. EOS/ESD Symp., 1995, pp. 43-61.
[28] Device modelling for the 1990s, Special Issue, Microelectronics Journal, Vol. 26, No. 2/3, March (1995), ISSN 0026-2692.[2]
[29] Stricker A., Mettler S., Wolf H., Mergens M., Wilkening W., Gieser H., Fichtner W., Characterization and Optimization of a Bipolar ESD-Device by Measurements and Simulations, to be published Proc. 20th EOS/ESD Symposium, (1998)
[30] Maex K., van Rossum M., Properties of Metal Silicides, INSPEC, London, U.K., ISBN 0 85296 859 0, 1995.
[31] Fornara P., Touret O., Poncet A., "Physical Models for 2D Numerical Simulation of New Titanium Salicide Process", Proc. ESSDERC Conf., 1996, pp. 145-148.
[32] Diaz C.H., Kang S.-M., Duvvury C., "Circuit-Level Electrothermal Simulation of Electrical Overstress Failures in Advanced MOS I/O Protection Devices", in IEEE Trans. on CAD of ICs and Circuits, Vol. 13, No. 4, (1994), pp.482-493.
[33] Russ C., Egger P., Gieser H., Irl S., "Compact Electro-Thermal Simulation of ESD-Protection Elements", Quality and Reliability Engineering International, Vol. 10, (1994), pp.335-340.
[34] Luchies J.R.M., de Kort C.G.C.M., Verweij J.F., "Fast Turn-On of an nMOS ESD Protection Transistor; Measurements and Simulations", Proc. EOS/ESD Symp., (1994), pp.266-272.
[35] Luchies J.R.M., Verhaege K., Kuper F., Mouthaan A., de Graaff H., "Fast Transient ESD Simulation of the nMOS Protection Transistor", Proc. 25th ESSDERC Conf., The Hague, The Netherlands, (1995), pp. 307-310.
[36] Tailliet F., Chante J.-P., "Characterisation of an n-p-n Structure under ESD Stress and Proposed Electrical Model", IEEE Trans. on Electron Devices, Vol. 37, No. 4, (1990), pp.1111-1120.
[37] Amerasekera A., Ramaswamy S., Chang M.-C., Duvvury C., "Modelling MOS Snapback and Parasitic Bipolar Action for Circuit-Level ESD and High Current Simulations", Proc. IRPS Symp., (1996), pp.318-326.
[38] Dutton R.W., "Bipolar Transistor Modelling of Avalanche Generation for Computer Circuit Simulation", IEEE Trans. on Electron Devices, Vol. 22 No. 6, (1975) pp.334-338.
[39] Hsu F.-C., Ko P.-K., Tam S., Hu C., Muller R.S., "An Analytical Breakdown for Short-Channel MOSFETs", IEEE Trans. on Electron Devices, Vol. 29, No. 11, Nov. (1982), pp.1735-1740.
[40] Scott D., Giles G., Hall J., "A Lumped Element Model for the Simulation of ESD Failures in Silicided Devices", Proc. EOS/ESD Symp., (1986), pp.41-47.
[41] Pinto-Guedes M., Chan P.C., "A Circuit Simulation Model for Bipolar-Induced Breakdown in MOSFETs", IEEE Trans. on Computer-Aided Design, Vol. 7, No. 2, (1988), pp.289-294.
[42] Skotnicki T., Merckel G., Denat C., "Triggering and Sustaining of Snapback in MOSFETs", Solid State Electronics,Vol.35,No.5, 1990, pp. 717-721.
[43] Krabbenborg B.H., de Graaff H.C., Mouthaan A.J., "Analytical Calculation of Avalanche and Thermal Snapback Points in Bipolar Transistors", Proc. ESSDERC Conf., Edinburgh, UK, (1994), pp.433-436.
[44] Krabbenborg B.H., Modelling and Simulation of Electrothermal Interaction in Bipolar Transistors, PhD thesis, Technical University Twente, The Netherlands, (1995), ISBN 90-9007589-5.
[45] Duvvury C., Amerasekera A., "Advanced CMOS Protection Device Trigger Mechanisms During CDM", Proc. EOS/ESD Symp., 1995, pp.162-174.
[46] Stricker A., Gloor D., Fichtner W., "Layout Optimisation of an ESD-Protection n-MOSFET by Simulation and Measurement", Proc. 17th EOS/ESD Symp., (1995), pp.205-211.
[47] Russ C., Kreisbeck J., Gieser H., Guggenmos X., Kanert W., "Electrothermal Device Simulation of a gg-nMOSt under HBM ESD Conditions", Proc. 6th ESREF Symposium, Arcachon, France, (1995), pp.141-146.
[48] Russ C., Verhaege K., Bock K., Roussel P.J.,Groeseneken G., Maes H.E., A compact model for the grounded-gate nMOS transistor under CDM ESD Strress, Jour. Electrostatics, 42 (1998),pp.351-381
[49] Bock K. and Hartnagel H.L., Fieldemitter-based ESD-Protection Circuits For High-Frequency Devices and IC's Proc. 14th EOS/ESD Symposium, (1992), p.168-174.
[50] Bock K., ESD Issues in Compound Semiconductor High-Frequency Devices and Circuits, Proc. 20th EOS/ESD Symposium, (1997), pp.1-12 (to be published in Microelectronics & Reliability 1998).

[2] because of the limited space for references here only the title of the special issue is given. The issue contents review and technical papers about process and device modelling.

PERGAMON

Microelectronics Reliability 38 (1998) 1009–1013

MICROELECTRONICS RELIABILITY

# Overview of the kinetics of the early stages of electromigration under low (= realistic) current density stress

J. Van Olmen[a], J.V. Manca[a], W. De Ceuninck[a], L. De Schepper[a], V. D'Haeger[a], A. Witvrouw[b], K. Maex[b]

[a] *Limburgs Universitair Centrum, Institute for Materials Research, Materials Physics Division, Wetenschapspark 1, B-3590 Diepenbeek (Belgium)*
[b] *IMEC, Kapeldreef 75, B-3001 Heverlee (Belgium)*

**Abstract**

The early stages of electromigration (EM) have been studied under realistic, i.e. low current densities ($j<0.5$ MA/cm$^2$) using a high resolution resistance measurement technique. Low current densities initiate EM and discard other masking mechanisms allowing an accurate observation of the EM kinetics, revealing fundamental features such as incubation time and subsequent linear resistance increase, important for modelling and life time extrapolation purposes. Current and temperature dependences are investigated and compared with the results obtained with high current density tests. For the first time it is shown that the processes responsible for the incubation time are reversible in nature. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Electromigration tests such as MTF (Median Time to Failure), WIJET (Wafer-level Isothermal Joule-heating Electromigration Test) and BEM (Breakdown Energy of Metal) use current density stress levels far beyond the normal operation conditions (see figure 1) and doubts can be expressed about the relevance of the obtained results to predict the life time under real operation conditions [1]. In this paper it will be pointed out that a more realistic approach to the study of the earliest stages of EM is possible through the use of EM tests with reduced current density. An important advantage of low current densities is that a negligible joule heating occurs, which eliminates thermal gradient induced failures and precipitation/dissolution effects. It has in fact been shown by Lloyd and Shatzkes [2] that the presence of thermal gradients can lead to the occurrence of failures at the ends of a stressed metallisation. And it is also known that even a small temperature change triggers reversible, physical phenomena such as precipitation/ dissolution of alloyed elements contributing to non-linear resistance changes [3, 4]. If the EM test is preceded by a pre-annealing step also other masking effects such as grain growth and defect relaxation can be eliminated and the resistance changes measured during the application

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00101-2

of the current stress are therefore related to EM only. This makes it possible to study accurately the earliest stages of EM which is of great importance for the modeling of the phenomenon. The measurement technique used to study the kinetics of the early stages of EM with low current densities will be presented below.

## 2. Measurement Techniques

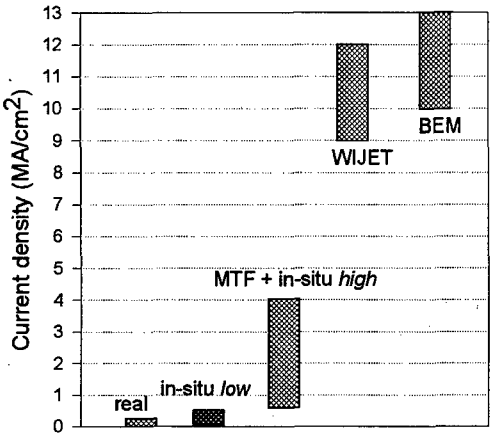



Fig. 1. Comparison of current density for various EM tests and real life conditions.

For the determination of the Median Time to Failure (MTF) the test temperature is typically between 175°C and 225°C and the current stress density $j$ is typically 2 MA/cm$^2$ or higher. The typical test time for this measurement technique is in the order of weeks to months. For the so-called 'fast techniques' such as WIJET and BEM the metal line is stressed until breakdown occurs and these techniques can there- fore be classified as destructive investigation techniques. The current density stress ($10^7$ A/cm$^2$ or more) is even higher than for MTF life tests, and the test time is reduced to seconds or minutes [5]. The current density stress is so high (typically 100 times or more than under real life conditions) that externally applied temperature stress is no longer required to heat up the interconnects ; the self heating (Joule heating) is sufficient to heat the tracks up to 175 °C or more.

In order to study electromigration on a more realistic base, current density stresses have been used that are comparable to real life conditions, i.e. j < 0.5 MA/cm$^2$. This has been possible with the high resolution resistance measurement technique, also called in-situ technique, developed a few years ago by IMO and IMEC, attaining a resolution of the order of 20 ppm for a metallisation with a TCR of 2000 ppm/°C, whereas the comparable resolution with a conventional method is in the order of 1000 ppm [6,7].

## 3. Low current density EM in Al and Al/Si lines

The samples that were used in this study were prepared and packaged at IMEC. All lines are polycrystalline in nature. Additional information regarding the samples is summarized in table 1.

Table 1 : sample information

| Type | metal | passivation | $T_{dep}$ | $d_{gr}$(μm) |
|---|---|---|---|---|
| A | Al | No | 60°C | 0.6 |
| B | Al 1%Si | No | 60°C | 0.6 |
| C | Al1%Si | SiN (380°C ) | 400°C | 3.8 |

Before performing low current density EM experiments, the samples are *pre-annealed* during several hours until no resistance variation is observed, in order to eliminate initial processes such as grain growth and defect relaxation. The elimination of these masking processes allows an accurate observation of the kinetics of the early stages of EM. When EM experiments are performed with high current densities (j>0.5 MA/cm$^2$) a linear time dependence of the resistance change is observed. However, if the metal lines are subjected to a low current density (j<0.5 MA/cm$^2$), the time dependence of the resistance is more complicated.

The result of such a low stress EM experiment is shown in figure 2 for non-passivated pure Al-lines and in figure 3 for passivated Al 1wt.%Si-metallizations. An initial incubation period is observed, followed by a linear resistance increase. However, at a certain moment, a deviation from this linear behaviour occurs. It is also observed that this deviation from linearity is different for passivated and non-passivated lines. The incubation period and linear resistance stage results are discussed briefly below.

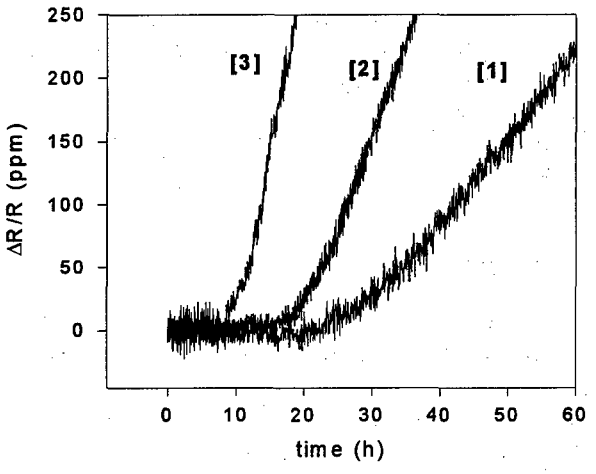



Fig. 2. Early resistance changes versus time for non-passivated pure Al-lines during low current density measurement at j = 0.25 MA/cm$^2$ and T=150°C[1], T=170°C[2],and T=210°C[3].

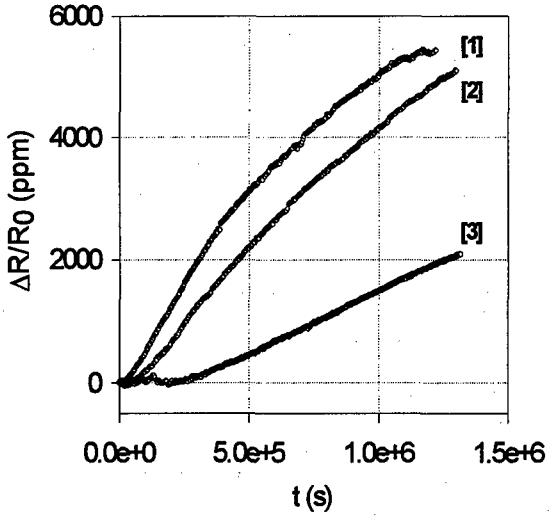



Fig. 3. Early resistance changes versus time for passivatedAl 1wt.%Si-metallizations during low current density stress at T=269°C and j = 0.13 MA/cm$^2$ [1], j = 0.09 MA/cm$^2$ [2] and j = 0.06 MA/cm$^2$ [3].

### 3.1. Linear Resistance increase

The linear stage of the resistance curves can be characterised by a *Rate of Resistance Change (RRC)*, given by the following expression :

$$RRC = \mathrm{d}(\Delta R/R_0)/\mathrm{dt}\,. \qquad (1)$$

The temperature and current density dependence of the RRC can be characterised by the following "Black-type" equation :

$$RRC = A \times j^n \times \exp(E_a/kT), \qquad (2)$$

where all symbols have their usual meaning. $A$ is a constant , depending only on the type of metallisation . The activation energy $E_a$ and the current density exponent $n$ were determined by performing experiments on *multiple* samples and a set of (j,T) pairs. The results for the three types of metallisation are summarised in table 2.

Table 2 : activation energy and current density exponent for linear stage

| Type | $E_a$(eV) | $n$ |
|---|---|---|
| A | 0.50 ± 0.02 | 2 ± 0.1 |
| B | 0.60 ± 0.02 | 2 ± 0.1 |
| C | 0.66 ± 0.02 | 2 ± 0.1 |

These results agree well with the activation energy and current density exponents as determined by high current density experiments. All these results are pointing in the direction of vacancy diffusion through the grain boundaries as the physical process underlying the linear resistance increase. Geometrical changes (void and hillock growth) are responsible for the linear resistance increase. This was shown in another paper by means of a resistivity/geometry decomposition method [8]. A possible explanation for the differences in activation energy for Al and Al1%Si could be the fact that vacancies in Al1%Si are bound to Si-atoms ,thereby effectively increasing the activation energy for diffusion. Also for the constant factor A important differences exist between the different types of metallisation. These differences can find their origin in e.g. diffences in initial stress state, in grain size,

initial concentration of Si atoms, and mechanical strength. Further analysis is needed to clear this point out.

The determination of activation energy and current density exponent is normally a time- and sample consuming task. We propose for the first time an alternative *quick* method to determine the

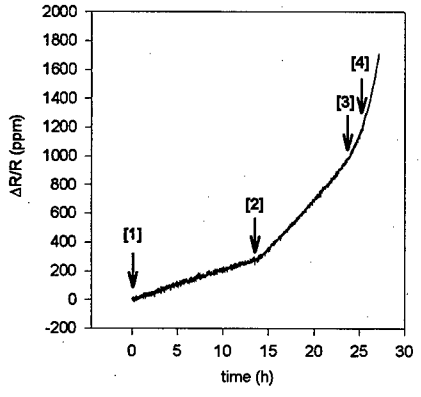



Fig. 4 Relative resistance change vs time for ***one*** pure Al stripe at T = 170°C and increasing current densities ($j_1$= 0.13 MA/cm$^2$,$j_2$ = 0.25 MA/cm$^2$, $j_3$= 0.38 MA/cm$^2$, $j_4$= 0.5 MA/cm$^2$).

current density dependence of the linear stage. With this method the resistance changes of just *one* sample are measured at high temperature during subsequently increasing EM current steps (low current densities). The low current densities are essential here to make sure that temperature rise due to Joule heating is negligible. The result of this procedure is shown in figure 4 for a non-passivated pure Al stripe. It is possible to determine the RRC of the linear stages during several hours, resulting in a linear relationship between the logarithm of the RRCs and the logarithm of the applied current densities, leading to a Black current density exponent n=2, as was also found with the traditional method (using more samples). Also for the determination of the activation energy this "one sample approach" can be used (i.e. switching the temperature at constant current density). This technique,also known as Overhauser's method, has since long been used for the determination of the activation energy of vacancy related processes [9]. When using Overhauser's method care should be taken that the time to switch from one temperature to another should be as small as possible. In our case this was achieved by using two oil baths (at different temperatures) and a specially designed sample holder that allows us to change quickly from one furnace to the other.

### *3.2 Incubation time*

Our experiments indicate that the temperature dependence of the incubation time obeys an Arrhenius relation with an activation energy in the order of 0.5eV~ 0.6 eV. It is also clear that the incubation time strongly depends on the current density stress. The incubation time for a passivated Al 1wt.%Si-metallization submitted to a current density j = 0.13 MA/cm$^2$ at T = 269°C is of the order of 9 hours, while for a current density of 0.06 MA/cm$^2$ the incubation time increases to 3 days. A lack of data points so far prevents us from determining current density exponents for this stage.

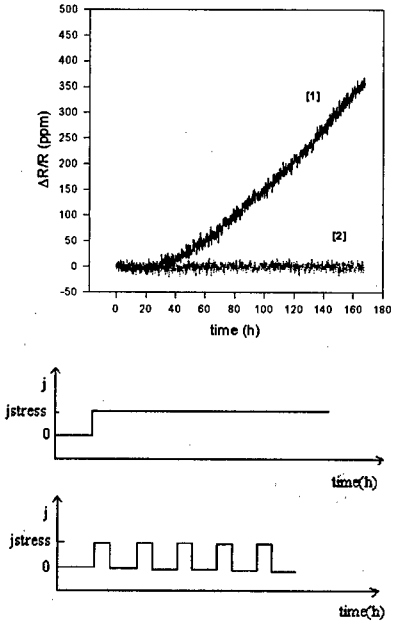



Fig. 5 a)Comparison of the relative resistance changes as a function of *cumulative* stress time for two pre-annealed stripes of non-passivated pure Al lines during an electromigration experiment at T=170°C. Stripe 1 is subjected to a continuous stress current (j=0.13 MA/cm$^2$), while stripe 2 is subjected to a succession of periods of 20 hours of stress followed by 48 hours of no stress; b) Current density profiles for both stripes.

With an appropriate experiment we have shown for the first time that the processes underlying the incubation time are of a reversible nature. Figure 5 shows the relative resistance changes as a function of time for a non-passivated pure Al line during continuous current stress (line 1) compared to the resistance changes on *another* line of the same type when subjected to non-continuous current stress (line 2). The incubation time of line 1 amounts to about 28 hours. The second line (for which the incubation time can also be expected to be around 28 hours) was then subjected to a current profile consisting of alternating periods of 20 hours of stressing followed by 48 hours of non-stressing. The curve for stripe 2 shown on the graph is the resistance change versus the ***cumulative*** stress time for this stripe, clearly showing that the processes underlying the incubation time are reversible in nature. All this implies that for real life operation conditions large incubation times are to be expected and the incubation time should be inserted for correct life time extrapolations. This is especially true for non- unidirectional current stress.

## 4. Conclusions

With the high resolution electrical measurement technique the early stages of electromigration can be studied underd realistic, i.e. low current densities ( $j < 0.5 MA/cm^2$ ). These low current densities initiate the EM process but discard other masking mechanisms, allowing for an accurate observation of the EM kinetics.

Early resistance changes are characterised by an initial incubation time, followed by a linear resistance increase. For the stage of linear resistance increase activation energies and current density exponents have been determined which correspond well with values determined in higher current density experiments ( $0.5\ MA/cm^2 < j < 5 MA/cm^2$). A new quick method, that requires the use of just one sample, has been introduced for the determination of the current density exponent $n$. All these results are pointing in the direction of vacancy diffusion through the grain boundaries as the physical process underlying the linear resistance increase. The incubation time also shows an Arrhenius-like temperature dependence. The incubation time also depends strongly on the current density, but results on current density exponents are inconclusive due to lack of data points so far. This will be studied in more detail in the future.

For the first time it is shown that the processes underlying the incubation time are reversible in nature. This is of great importance for modelling, but has also implications for life time extrapolations, since for real life operation conditions very large incubation times are to be expected.

## References

[1] J.R. Lloyd, Mat. Res. Soc. Symp. Proc., Vol. 265, p. 177-186, (1992).

[2] J.R. Lloyd, M. Shatzkes and D.C. Challener, Proc. of the 8th IEEE International Reliability Physics Symposium, pp. 216-225, (1988).

[3] A. Scorzoni, I. De Munari, H. Stulens, V. D'Haeger, J. Appl. Phys. 80 (1), 1 July 1996, pp. 143-150, (1996).

[4] W.A. De Ceuninck, V. D'Haeger, J. Van Olmen, A. Witvrouw, K. Maex and L. De Schepper, accepted for publication in Microelectronics and Reliability.

[5] A. Scorzoni, B. Neri, C. Caprile, F. Fantini, Mat. Sci. Reports, Vol. 7, p. 143-220, 1991

[6] European patent (EP 0395 149 A1) 'Method and device for accelerated determining of ageing of one or more elements with an electromagnetic ageing parameter', W.A. De Ceuninck, L. De Schepper, L. Stals and J. Roggen.

[7] W. De Ceuninck, J. Manca, V. D'Haeger, L. De Schepper and L.M. Stals, ESREF 95

[8] V. D'Haeger, W. De Ceuninck, G. Knuyt, L. De Schepper and L.M. Stals, MRS Proc. 1996, Vol. 428, p. 133

[9] A. C. Damask and G.J. Dienes, Point Defects in Metals, London, Gordon and Breach, 1971

PERGAMON

Microelectronics Reliability 38 (1998) 1015–1020

MICROELECTRONICS RELIABILITY

# Effects of alloying elements on electromigration

R. Spolenak, O. Kraft, E. Arzt

*Max-Planck-Institut für Metallforschung, 70174 Stuttgart, Germany*

**Abstract**

Alloying elements, such as Cu, are used to reduce electromigration damage in miniaturized Al conductor lines. Nevertheless a thorough understanding of the fundamental mechanisms governing alloying effects has not been achieved yet. We present *in situ* drift experiments on pure Al and the alloys Al-Cu and Al-Mg and interpret them on the basis of activation energies. A mechanistic model is developed in order to account for an increase in activation energies for Al drift in the alloys in comparison to the pure metal. Good agreement is found between the model and the experiment. © 1998 Elsevier Science Ltd. All rights reserved.

## Introduction

Electromigration still is an important reliability concern in microelectronics. Copper as an alloying element for aluminum interconnects has been found to strongly improve their electromigration resistance [1-5]. However fundamental understanding of the underlying mechanisms has not been achieved yet. These mechanisms are very important, especially when lifetime data measured under accelerating conditions are extrapolated to use conditions.

In order to obtain more insight into alloying effects on electromigration we carried out a comparative investigation on Al-Cu and Al-Mg. The latter is also known to improve the electromigration resistance [6,7], but shows a completely different microstructure. The solubility of Mg in Al is about 6 wt% at 500 K whereas that of Cu is only about 0.2 wt%. As a result Al-Cu is a precipitate hardening system, whereas Al-Mg is a solid solution hardener.

The atomic drift velocities caused by electromigration are usually described by:

$$v = \frac{D}{kT} eZ^* \rho j \tag{1}$$

where $D$ is the diffusion constant for the dominating diffusion process, $e$ $Z^*$ the effective charge, ρ the electrical resistivity and j the current density.

An experiment in which this drift velocity can be directly observed is the so called "Blech" experiment [8]: A stripe of metal, in this case Al or Al-alloys, is deposited on a continuous line of significantly lower conductivity. Thus when current is applied, it flows through the metal rather than the support structure. As the metal stripe is confined geometrically, the drift of atoms is directly reflected in the displacement of the cathode edge of the stripe.

If the stripe length becomes short, the atoms encounter the effect of a backdriving force due to the build-up of mechanical stress. It can even compensate the electromigration force at a critical

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00110-3

length $l_c$ so that no stripe drift occurs at all. The critical condition is given by:

$$jl_c = \frac{\Omega\Delta\sigma}{eZ^*\rho} = const. \quad (2)$$

where $\Delta\sigma$ is the maximum possible stress difference between the ends of the stripe and $\Omega$ is the atomic volume.

In Al-Cu alloys the additional observation was made that Al drift does occur, only after the critical length is depleted of Cu precipitates, which is observed as an incubation time in drift experiments [1].This is due to the fact that Cu strongly inhibits the self diffusion of Al [9].

## Experimental

We have performed drift experiments on pure Al, Al2wt.%Cu and Al3wt.%Mg. Aluminum alloy stripes were sputter deposited on TiN and structured by a lift-off technique. All stripes were 500 nm thick and 5 μm wide. The experiments were carried out *in situ* in an SEM. The stripe drift was quantified using secondary and backscattered electron images. Simultaneously concentration profiles of the alloying elements were measured using an EDX detector. In this way the drift of the matrix element as well as the alloying element could be quantified independently. Each alloy was tested at various temperatures ranging from 175 °C to 275 °C to allow the determination of activation energies that are indispensable for an extrapolation to service conditions and provide insight in the basic governing mechanisms. The current density was set to be 1 MA/cm$^2$. To stabilize the microstructure a heat treatment of 12 hr at 225 °C in air was carried out prior to testing. The grain size was measured by taking a series of three Focused Ion Beam (FIB) images at different tilt angles and a subsequent image analysis.

## Results

### *Drift velocities*

In Fig. 1 the displacement of the cathode end of the stripe is plotted against time. Pure Al starts to drift immediately after an electrical current is applied to the structure. In the cases of the alloys Al-Cu and Al-Mg, one can observe an incubation time in which practically no macroscopic drift is observable. The incubation time for Al-Cu can be correlated to depletion of copper over a critical length for pure Al of about 10 μm at 1 MA/cm$^2$, as has been determined by the EDX measurements.

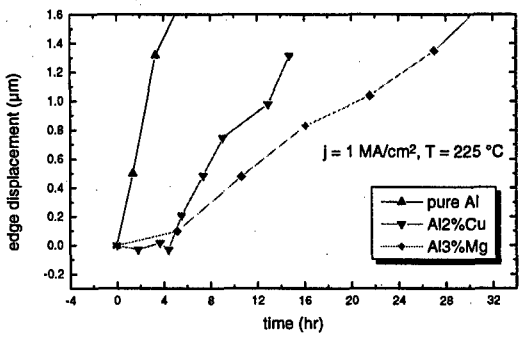



Fig.1. Comparison of the edge displacements for Al, Al-Cu and Al-Mg as a function of time

Transport of Mg atoms in Al-Mg could be observed, but the resolution of the analysis system was not sufficient to make a statement about a Mg depleted zone.

Further it can be observed that even after the incubation time the alloys do not behave like the pure metal. The drift velocities are still reduced in comparison to the pure metal. In Al-Mg the effect appears to be even stronger than in Al-Cu, especially considering the smaller grain size of Al-Mg (Al: 250 nm, Al-Cu: 250 nm, Al-Mg: 110 nm).

### *Activation energies*

The results of the temperature-dependent drift experiments are shown in the form of an Arrhenius plot in Fig. 2. The drift velocities of Al in the alloys were obtained as the fitted slopes of the curves in Fig. 1 after the incubation time. The drift velocity of Al in Al-Mg was divided by a factor of 2.5 to account for the difference in grain size. For the drift of Cu the concentration profiles determined by EDX measurements were analyzed, and the drift velocity was determined for the center of mass of Cu during the incubation time for Al [10].

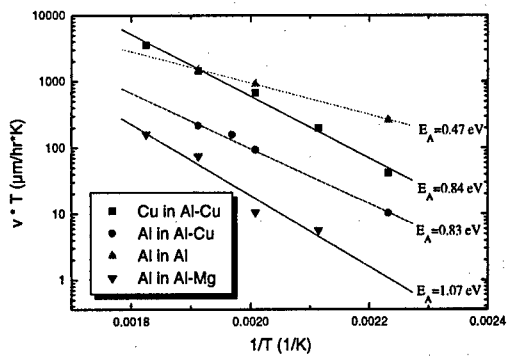



Fig. 2.Drift velocities of Al after the incubation time in various alloys as a function of inverse temperature

The activation energy of pure Al drift corresponds to literature data. In the case of alloys, however, the drift of Al after the incubation time does not behave as in pure Al. It shows higher activation energies of about 0.8 eV in Al-Cu and 1 eV in Al-Mg. In Al-Cu the activation energy of the Cu drift coincides with that for the Al drift in the alloy and is in agreement with Cu diffusion in Al-Cu grain-boundaries [11,12].

## A model for electromigration describing sink and source action

The model is described schematically in Fig. 3: Without current, the stripe contains an equilibrium concentration $c_0$ of atoms. When the electrical current is turned on, the electromigration force acts on the atoms and a transport to the anode side of the stripe takes place. There they change the local atomic concentration giving rise to two effects: First a concentration gradient for the atoms is established along the stripe; this results in a backdriving force counteracting the electromigration force and thus reducing the net flux of arriving atoms at the anode end. Second, as the atomic concentration rises, an "osmotic" pressure is exerted on the sinks for atoms. The sink action allows for the formation of hillocks by absorbing atoms. In the special case of Al and Al-alloys it is appropriate to describe the sinks as grain boundary dislocations, which climb as they emit or absorb atoms [13]. In a mechanically constrained system this process also causes the build-up of mechanical stress.

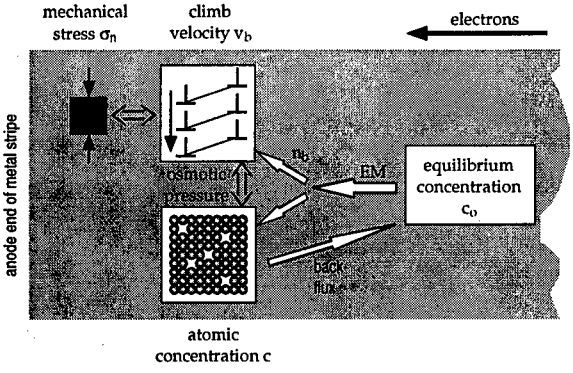



Fig. 3. Schematic description of the new model

These stresses can counteract the osmotic pressure and thus slow down and finally stop the dislocation motion.

The details of the model are described elsewhere [14]. Here, we cite only the term that governs the dislocation motion.

$$v_b = Mb_b\left[\sigma_n - \frac{kT}{\Omega}\ln\left(\frac{1-c}{1-c_0}\right)\right] \tag{3}$$

where $b_b$ is the Burgers vector of the dislocation. The climb velocity $v_b$ is determined by the driving force (the expression in brackets) and the dislocation mobility $M$. The force is the balance between the normal stress $\sigma_n$ on the dislocations and the osmotic pressure. The latter is determined by the actual atomic concentration $c$ compared to the equilibrium concentration $c_0$, the atomic volume $\Omega$ and the thermal energy $kT$.

It is the mobility term that accounts for alloying effects. For a pure metal, the mobility of dislocations is given by:

$$M = \frac{D_b b_b}{kT} \tag{4}$$

where $D_b$ is the grain boundary diffusivity of the matrix.

In an alloy, solute atoms redistribute in the stress field of dislocations and have to be moved, when the dislocation climbs [15]. Therefore a process that in the pure metal is controlled by grain boundary diffusion, is now controlled by volume diffusion of the solute as follows:

$$M = \frac{D_s \Omega}{\beta k T b_b^2 c_s} \quad (5)$$

$D_s$ is the volume diffusivity of the solute, $c_s$ is its atomic concentration and β is a dimensionless parameter characterizing the interaction between the dislocation and the solute. It depends on the nature of the solute atom and usually varies between 3 and 20.

In pure metals the transport of atoms is the rate limiting step, but once alloying elements are added the climb velocity of the dislocations, which determines the effectiveness of the sinks, becomes rate limiting. This way, the model can account for alloying effects, but it is not able to explain a critical stripe length. In order to account for a "Blech" effect, a critical, threshold, concentration $c_c$ has to be established, below which the actual stripe drift and thus the dislocation mobility is zero. This corresponds to the real situation of a passivation that has to be fractured before stripe drift occurs (the natural passivation of Al, for example).

## Comparison between experiment and model

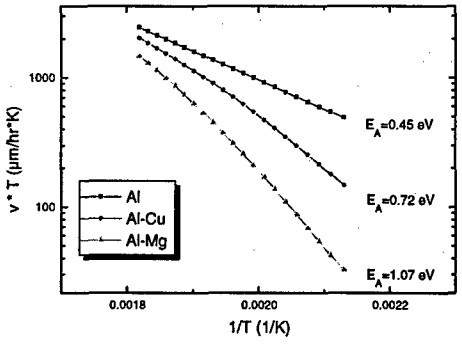



Fig. 4. Results of the model for different alloys

For the comparison of the results of the model and the experimental results actual numbers have to introduced in the model. Nearly all of the values can be taken from literature. Only the grain boundary diffusivity of aluminum is difficult to determine as there is no easily available radioactive tracer. Therefore the experimental results for the pure aluminum drift are used. Another unknown is the interaction parameter β, which is chosen to be 10. The other input data are summarized in Table 1. Instead of using separate values for the effective charge $Z^*$ and ρ, the specific electrical resistivity the value of their product is used, because it is temperature independent [16].

Table 1
Input data for the model

| parameter | value |
|---|---|
| $D_{0gb,Al}\,\delta$ | $6\ 10^{-17}$ m²/s (exp.) |
| $E_{gb,Al}$ | 0.47 eV (exp.) |
| $D_{0V,Cu}$ | $6.4\ 10^{-5}$ m²/s [17] |
| $E_{V,Cu}$ | 1.4 eV [17] |
| $D_{0V,Mg}$ | $5\ 10^{-5}$ m²/s [18] |
| $E_{V,Mg}$ | 1.31 eV [18] |
| $Z^*\rho$ | -50 µΩcm |
| β | 10 |
| $c_{Cu}$ | $e^{-3421/T}$ 100 at% [19,20] |
| $c_{Mg}$ | 3 at% |
| $j$ | 1 MA/cm² |
| $l_c$ | 10 µm |
| $L$ | 30 µm |

Fig. 4 shows results of the in the form of an Arrhenius plot. As the calculated drift velocities of the alloys are plotted against the reciprocal temperature, apparent activation energies can be determined from the slope. Qualitatively the model shows the correct results when compared to the experiment of Fig. 2: Al-Mg shows the strongest alloying effect.

Quantitatively the model is not fully predictive. The relative alloying effect between the two alloys is correct, but it should be stronger in both cases in comparison to pure Al. This fact, however, can easily be explained: The model describes a non-local situation, i.e. the situation only at the site, of hillock formation. Non-local effects like the backdriving force are included, but not the concentration variation of the alloying element along the line that lead to incubation times. Therefore the actual effective stripe length after the incubation time is not equivalent to the geometrical stripe length, on which the model is based. As the effective stripe length, in which Al is allowed to diffuse "freely", is shorter, additional retarding effects are present. They are responsible for the discrepancy between the experiment and the model.

If activation energies are considered, the situation is different. For a choice of the interaction parameter β of 10 the calculated activation energies

agree quite well with the experimental results. Certainly the value of the activation energies can be changed by the choice of the interaction parameter. However, the activation energies for both Al-Cu and Al-Mg fit for one choice of this parameter. This is a strong indication that the fundamental alloying effects are described correctly.

Furthermore one can note that in Fig. 4 the curves in the Arrhenius plot are not linear, which would be the case if one had only one activation energy. This is a result of the model that incorporates different mechanisms with different activation energies. It has significant effects on lifetime predictions, because one activation energy determined in a narrow temperature range will not suffice in describing the system far outside this temperature range.

## Lifetime predictions

In order to take the step from a theoretical model to actual lifetime predictions, certain assumptions have to be made. Let us consider a microprocessor with several levels of interconnects. The vertical connections are made by vias that act as diffusion barriers. Then, assuming that the lines have polycrystalline grain structure, the longest stripes that carry the highest current densities are most likely to fail first. Therefore, in order to predict the lifetime of all interconnects, it will suffice to consider the weakest segment in the chain. Furthermore we assume the effective diameter of the via to be $d_{via}$ and the length of the segment $L$. Then the actual time to failure ($TTF$), i.e. the time until the whole via diameter is depleted of Al, becomes:

$$TTF = t_{inc}(T, j) + \frac{d_{via}}{v(L, l_c, T, j)} \qquad (6)$$

where $t_{inc}$ is the incubation time for stripe drift. It is primarily dependent on the diffusion process of the alloying element and can therefore be described by one activation energy, namely the one for the diffusion of the solute in Al grain boundaries. It can be determined by the experiment described above.

Once the incubation time is over, actual stripe depletion occurs. Then the additional time for failure to take place is the time needed for the cathode stripe edge to drift the distance of the via diameter $d_{via}$. This can be calculated by making use of the model. The process is controlled by Al self diffusion in grain boundaries as well as the diffusion of the solute in the Al lattice.

Concluding it can be said that the actual time to failure is composed of several different mechanisms. All of them have to be included to allow for an exact extrapolation to service conditions.

## Comparison of the alloys

From the results of drift experiments, Al-Mg seems to be a very good alternative to Al-Cu. Apart from already showing better electromigration resistance at testing temperatures, it additionally has higher activation energies that should make the improvement at service conditions even greater. However, if the alloying effect is normalized by the nominal concentration the situation seems to be different as many more Mg atoms are present in the alloys investigated. On the other hand it has been shown that Mg shows the tendency to segregate to the surface and so the effective concentration is reduced [10]. Overall Mg seems to be at least as effective as Cu.

Technologically, Al-Mg is not very relevant because of the high solubility of Mg. The increase of the specific resistivity pure Al amounts to 0.55 μΩcm per wt% of Mg [21] in solution. Thus the resistivity is much too high.

## Summary and conclusions

A comparison of the electromigration behavior of Al, Al-Cu and Al-Mg has been performed:

Al-Cu as well as Al-Mg show an incubation time before Al drift occurs. The incubation time can be attributed to the redistribution of the alloying elements, which is in agreement with other observations [1].

However, even after the incubation times alloying effects are noticeable. They lead to a reduction of absolute drift velocities in comparison to the pure metal as well as significantly higher activation energies.

The increase of activation energies is interpreted on the basis of a newly introduced sink and source model. The alloying effects are included as solute drag effects on dislocations acting as sinks and sources. The quantitative agreement between the

actual activation energies and the calculated ones is very good.

From this, a simple lifetime prediction for multilevel structures is deduced. As can be seen from the results of the model, lifetime predictions based on only one activation energy can be misleading and are always too conservative.

The nonlinearity of drift velocities in the Arrhenius plot for Al-Cu may be an explanation for strongly varying results for activation energies in the literature: The result is strongly dependent on the temperature regime investigated.

The quantitative agreement of drift velocities of the alloys compared to the pure metal between the model and the experiment is not yet satisfactory, which can be attributed to non-local alloying effects: The redistribution of alloying elements along the stripe leads to an effective stripe length that is different from the geometrical one. Further extension of the model to include these effects is currently in progress.

## References

[1] C.-K. Hu, M. B. Small, and P. S. Ho, J. Appl. Phys. **74** (2), 969-978 (1993).
[2] C.-K. Hu, K. Y. Lee, K. L. Lee, Jr. Cabral, C., E. G. Colgan, and C. Stanis, J. Electrochem. Soc. **143** (3), 1001-1006 (1996).
[3] C. Kim and J.W. Morris, Jr., J. Appl. Phys. **72**, 1837 (1992).
[4] I. Ames, F. M. d'Heurle, and R. E. Horstmann, IBM J. Res. Dev. **14** (July), 461-463 (1970).
[5] I. A. Blech, J.Appl.Phys. **48** (2), 473-477 (1977).
[6] A. Gangulee and F. M. d'Heurle, Thin Solid Films **25**, 317-325 (1975).
[7] M. Akiya and M. Ishii, J. Electrochem. Soc. **141** (10), 2857-2859 (1994).
[8] I. A. Blech, J. Appl. Phys. **47** (4), 1203-1208 (1976).
[9] R. Rosenberg, J. Vac. Sci. Tech. **9** (1), 263-270 (1972).
[10] R. Spolenak, J. Mason, O. Kraft, and E. Arzt in "MRS Spring Meeting", San Francisco, CA, Materials Research Society, to be published, 1998
[11] A. D. Brailsford and H. B Aaron, Jour. Appl. Phys. **40**, 1702 (1969).
[12] J. Goldman, H. I. Aaronson, and H. B. Aaron, Met. Trans. **1**, 1805 (1970).
[13] J. P. Hirth and J. Lothe, Theory of Dislocations McGraw Hill, 1968, ch. 15, p. 612
[14] R. Spolenak, O. Kraft, W. D. Nix, and E. Arzt in "Stress induced phenomena in metallization", Tokyo, Japan, AIP Conference Proceedings, 418, 147-158, 1997
[15] E. Arzt, F. Ashby, and R. A. Verrall, Acta metall. **31** (1983).
[16] J. P. Dekker, A. Lodder, and J. van Eck, Phys. Rev. B **56** (12), 167 (1997).
[17] N. L. Peterson and S. J. Rothman, Phys. Rev. B **1** (8), 3264-3273 (1970).
[18] Y. Minamino and T. Yamane, J. Mat. Sci. **38**, 2677 (1983).
[19] J. L. Murray, Int. Met. Rev. **130** (5), 211-233 (1985).
[20] P. R. Besser, "X-ray determination of thermal strains and stresses in thin aluminum films and lines," , Stanford University, 1993.
[21] C. Kammer, Aluminium-Taschenbuch, 14 Aluminium-Verlag, Düsseldorf, 1995,p. 198

PERGAMON

Microelectronics Reliability 38 (1998) 1021–1027

MICROELECTRONICS RELIABILITY

# A comparison between normally and highly accelerated electromigration tests

S. Foley [a], A. Scorzoni [b], R. Balboni[b], M. Impronta [b],I. De Munari [c], A. Mathewson [a] and F. Fantini [c]

[a] NMRC, University College Cork, Lee Maltings, Prospect Row, Cork, Ireland

[b] CNR - Istituto LAMEL, via P. Gobetti 101, 40129 Bologna, Italy

[c] Centro MTI, Università di Parma, v.le delle Scienze, 43100 Parma, Italy

**Abstract**

Normally and highly accelerated electromigration tests on Al-Cu lines of different widths are compared. It is shown that the use of the Black equation gives different extrapolated results depending on the range of stress conditions considered. It is concluded that for extrapolations it is safe to use the Black equation only in the case of normally accelerated stress conditions (temperatures in the range 150-240°C and current densities between 1 and 2 MA/cm$^2$) and when the extracted value for the current density acceleration factor is 2.
© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Conventional techniques for electromigration (EM) characterization often require very long measurement times [1] and for this reason, fast methods are also employed to improve test throughput. A decreased measurement time can be achieved by using very highly accelerated stress conditions. The first example of fast techniques is the Standard Wafer-level EM Acceleration Test (SWEAT) [2], proposed in 1985 as a very rapid characterization technique for EM. This technique is able to give results in less than 15 seconds thanks to a special test structure able to maximize temperature gradients. The main feature of the method is that the temperature increase, necessary to accelerate the test, is obtained by means of Joule heating only. A number of variations of the SWEAT technique have also been introduced in the last decade. The isothermal test, introduced in 1987 [3] and currently adopted by Sandia Technologies in their Wafer Level Reliability EM test [4], uses a standard, straight line test pattern in order to reduce the problem of large thermal gradients. Giroux et al. [5] proposed the Constant Acceleration Factor EM test (CAFE) which is a Joule-heated SWEAT test using a standard JEDEC structure and slightly less accelerated test conditions (maximum temperature was 300°C).

In this paper an attempt has been made to answer the question: "how accelerated can an EM test be?". Conventional and relatively fast EM tests performed on the same set of samples will be described and the results of EM parameter extractions will be compared. Finally, conclusions will be drawn on the consequences of using EM parameters extracted from fast methods.

## 2. Experimental and data analysis

EM life tests were performed on Al-Cu lines with a TiN/Ti barrier and a TiN ARC. Tests were

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00121-8

mainly performed on 5 μm wide lines, but a limited number of tests were also performed on 3 μm wide lines. All 5 μm wide specimens came from the same lot, while 3 μm wide specimens were drawn from two different lots. Normally accelerated tests have been performed at average line temperatures of 175, 210 and 240°C and current densities of 2 and 3 MA/cm$^2$. Highly accelerated tests have been performed on 5 μm wide lines only, at a constant line temperature of 250°C and current densities from 4.5 to 6.5 MA/cm$^2$. The high values of current density used in this case induce a self-heating (due to Joule effect) much greater than that observed in the normally accelerated case, but smaller than in the case of commonly used rapid Joule-heated tests. Table 1 and Table 2 list the average thermal resistances and Joule self-heating of 5 μm wide and 3μm wide specimens respectively.

Table 1
Average thermal resistances and Joule self-heating of 5 μm wide specimens.

| stress condition | | average $R_{th}$ | average $\Delta T$ |
|---|---|---|---|
| °C | MA/cm$^2$ | °C/W | °C |
| 240 | 2 | 391.4 | 17.6 |
| 175 | 3 | 416.3 | 37.8 |
| 210 | 3 | 426.1 | 41.3 |
| 240 | 3 | 404.8 | 42.2 |
| 250 | 4.5 | 417.6 | 99.3 |
| 250 | 5.2 | 418.9 | 134.5 |
| 250 | 6 | 417.2 | 176.2 |
| 250 | 6.5 | 418.5 | 205.3 |

Table 2
Average thermal resistances and Joule self-heating of 3 μm wide specimens.

| stress condition | | average $R_{th}$ | average $\Delta T$ |
|---|---|---|---|
| °C | MA/cm$^2$ | °C/W | °C |
| 175 | 3 | 565.3 | 31.2 |
| 240 | 2 | 548.8 | 14.9 |
| 240 | 3 | 512.6 | 32.5 |

In normally accelerated tests the maximum self-heating was approximately 40°C in the case of 5 μm wide lines and approximately 33°C in the case of 3 μm wide lines. In highly accelerated tests on 5 μm wide lines the average self-heating ranges, depending on the stress conditions, between 100°C and 205°C. In all of the tests performed, the stress temperature of each and every specimen was accurately measured and successively taken into account in the data analysis. A failure criterion of 10% change in line resistance was adopted for the life-time analysis.

Normally accelerated tests followed the ASTM test procedure [6]. However, the ASTM standard has been "revisited" during this work and some weak points (like the TCR measurement) have been improved by taking other existing standards into account. *Constant current* was forced through the metal lines during EM tests. Moreover, for each stress condition the *same initial current density* was applied to all specimens under test by finely adjusting the current of each specimen using a procedure described elsewhere [7]. Obviously, the current density will change with time due to changes of the cross section.

The highly accelerated tests were simply isocurrent tests, without fine current adjustment.

To analyze the collected data, a fitting procedure was adopted, based on the Black equation (see Eq.1) for the Median Time to Failure

$$\mathrm{MTF} = Aj^{-n} \exp(E_a / k_B T) \qquad (1)$$

Here $j$ is the current density, $E_a$ is the activation energy of the relevant EM process, $n$ is the current density acceleration factor, $k_B$ is Boltzmann's constant, $T$ is the absolute temperature and $A$ is a suitable empirical constant. A theoretical value of 2 for the parameter $n$ has been demonstrated [8]. Values of $n > 2$ indicate abnormally accelerated stress conditions.

A lognormal distribution for Times to Failure (TTF) was assumed [1]. For all of the data sets a common failure mechanism was postulated (and successively verified by failure analysis). This implies that the lognormal standard deviation $\sigma$ is the same for all the data sets for all stress conditions.

The fitting parameters were $E_a$, $n$, the $\mathrm{MTF}_0$ at "operating" conditions $T_0 = 125$°C and $j_0 = 0.5$ MA/cm$^2$ (in place of the pre-exponential constant $A$) and $\sigma$. The use of this fitting procedure allowed us to perform a *design of experiments* in such a way as to decrease the number of tests with respect to the traditional "3 $T$ — 3 $j$" test (3 current densities and 3 temperatures)

## 3. Discussion of results

### 3.1 Results on 5 μm wide lines

Normally accelerated life test measurements on 5 μm wide lines have been compared in Figure 1 to the obtained best fits.

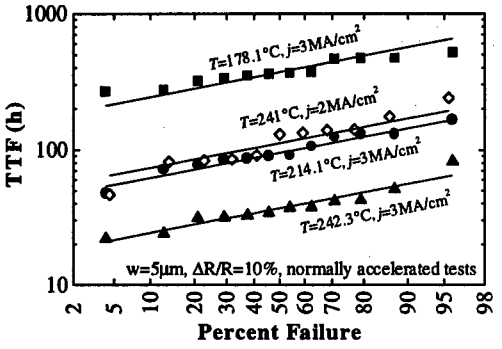



Figure 1. Results of normally accelerated tests (5 μm wide specimens). Solid lines are the result of the fitting procedure described in the text (the calculated failure rate at $10^5$ h is less than 0.1 FIT).

The extracted parameters are listed in the first column of Table 3.

Table 3
Comparison of the parameters extracted in different ranges of stress conditions for 5 μm wide lines.

| | Normally accelerated only | all together | highly accelerated only |
|---|---|---|---|
| $E_a$ (eV) | 0.72 | 0.76 | - |
| $n$ | 2.63 | 3.07 | 3.31 |
| σ | 0.32 | 0.33 | 0.36 |
| $MTF_0$ (h) | 500,327 | 1,202,604 | 1,706,713 dependent on $E_a$ (see text) |

The obtained activation energy falls in the well-accepted range for wide Al-Cu lines and the standard deviation of life-times is very small, thus giving a first confirmation to the hypothesis of a common failure mechanism for all test conditions. However, the value of $n$ is higher than the theoretical value of 2 which could indicate the possible presence of Joule heating-induced temperature gradients. This point of view is supported by the relatively high values of $\Delta T$ corresponding to 3 MA/cm$^2$ in Table 1 (about 40°C).

Postulating the validity of the Black equation even at high levels of stress, the best fit of the highly accelerated experimental data (see Figure 2) gives the parameters listed in the third column of Table 3.

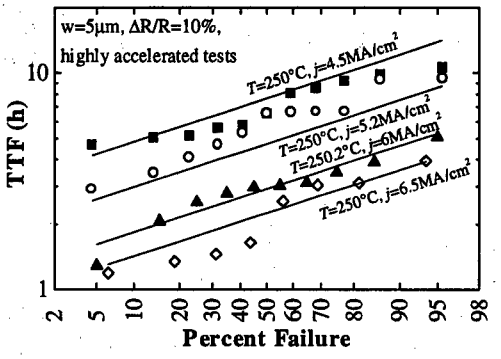



Figure 2. Results of highly accelerated tests (5 μm wide specimens). Solid lines are the result of the fitting procedure.

For this calculation the value of activation energy $E_a$ extracted from normally accelerated tests was adopted , since all the highly accelerated tests are performed at the same temperature (250°C). Therefore, the significant difference between the life-times at operating conditions is mainly due to the difference in the current density exponent $n$. It should also be noted that the disagreement between extracted parameters is noticeable notwithstanding the "relatively low" temperature and current densities adopted for our highly accelerated tests. In fact SWEAT and similar tests are usually performed at temperatures greater than 300°C and current densities of the order of (or greater than) 10 MA/cm$^2$.

A failure analysis was performed on failed specimens in order to detect the possible existence of different failure modes, depending on the stress conditions. For all stress conditions the failures were generally caused by extended voids together with a number of line-edge voids. Normally, extended voids were coupled with hillocks downstream (as illustrated in Figure 3).

Despite the fact that the failure analysis was performed on specimens that reached the same failure criterion (relative resistance increase of 20%), the average size, the number and the degree of void clustering of line-edge voids increases as the stress condition worsens.

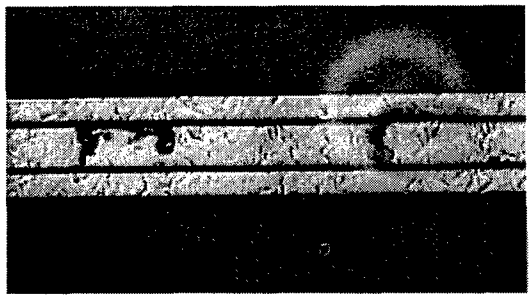


Figure 3. Line-edge, extended void and hillock in a specimen stressed at 240°C and 2 MA/cm$^2$. The micrograph shows the biggest voids found in the line.

For the higher stress conditions the amount of void clustering is such that an increasing influence of temperature gradients at defects can be expected (see Figure 4). It is well accepted that this problem normally induces an increase of the current density acceleration factor, and this was observed in these measurements.

No evidence of line-end failures was found, indicating that thermal gradient-induced failures at line-ends are absent in these tests.

This failure analysis demonstrates that worsening the stress conditions in a continuous fashion leads to a continuous change in the failure mode (increasing size, number and degree of clustering of line-edge voids). This in turn leads to an increased influence of temperature gradients at the void clusters. However, despite this continuous change of the failure mode, all the data sets collected at different stress conditions show approximately the same standard deviation (excluding the very scattered data collected at the worst stress condition, 250°C and 6.5MA/cm$^2$).



Figure 4. Line-edge voids in a specimen stressed at 250°C and 4.5 MA/cm2. The micrograph shows a typical cluster of voids.

Once it was verified that the standard deviation of these tests was approximately the same, an attempt was made to fit all data sets at all stress conditions with only one set of parameters (see Figure 5). This is what an undiscerning reliability engineer could have done in the case of the absence of a failure analysis.

The results of the fitting procedure are listed in the second column of Table 2. Even if the fit of Figure 5(a) looks acceptable, the calculated fitting error is 50% worse than in Figure 1.

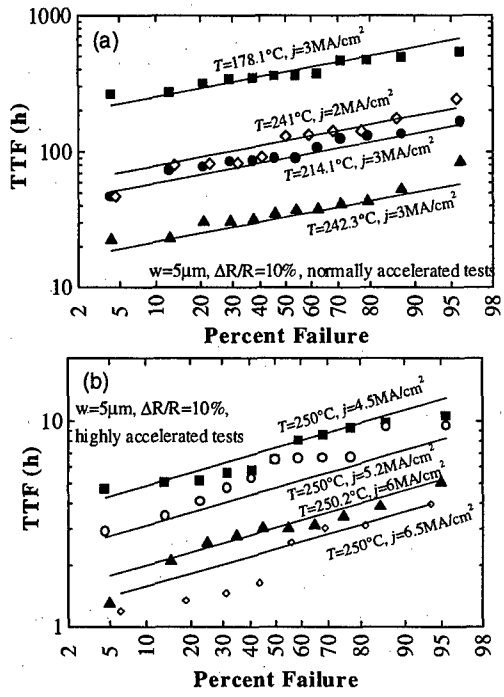



Figure 5. Fitting of all data with only one set of parameters (5 μm wide specimens). (a) Normally accelerated tests; (b) highly accelerated tests.

From Table 3 it turns out that the parameters extracted using the Black equation depend on the stress range, even in this case where the real temperature of the specimens is correctly taken into account.

### *3.2 Results on 3 μm wide lines*

Lifetime results of normally accelerated tests on 3 μm wide lines are shown in Figure 6. The solid lines shown in Figure 6 are best fits to the experimental data. The parameters directly extracted are $E_a = 0.91$ eV, $n = 3.5$, $\sigma = 0.72$, $MTF_0(T_0=125°C, j_0=0.5$ MA/cm$^2) = 28{,}769{,}752$ h.

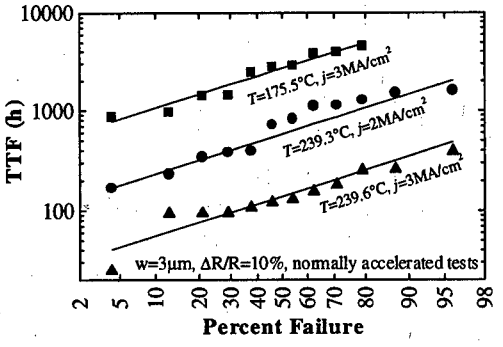



Figure 6. Results of normally accelerated tests on 3 μm wide specimens.

The parameters $E_a$, $n$ and $\sigma$ are higher than the values expected for polycrystalline lines. In particular, the values of $E_a$ and $\sigma$ are typical of cluster-bamboo structures, while a value of $n = 3.5$ could be associated with large thermal gradients at line ends or at defect locations. However, since it is known that the samples come from two different lots, a simpler explanation could be that the two lots have different median times to failure at the various stress conditions used, thus inducing an artifact in the extracted parameters. A closer look at Figure 6 reveals that the data at 175°C, 3 MA/cm$^2$ and 240°C, 2 MA/cm$^2$ both seem to follow a bimodal distribution, while the specimens which were subjected to the most accelerated stress condition seem to belong to one population.

Optical failure analysis confirmed this observation. All of the specimens stressed at 240°C and 3 MA/cm$^2$ had very similar failure modes (line break by a single void and several edge voids, see Figure 7). However, both sets of specimens stressed at 240°C and 2 MA/cm$^2$ and 175°C and 3 MA/cm$^2$ could be subdivided into two populations, a weak one with a few edge voids and a strong one which experienced the creation of many edge voids before the catastrophic failure. Moreover, only two line-end voids were found and this excludes systematic EM failures due to thermal gradients at line ends. Instead, failures were found at random locations both by extended voids and by slit-like voids (see Figure 8). The latter indicate that some elements of the population could feature some bamboo sections. An SEM analysis was performed on ion-milled specimens that failed due to slit-like voiding. As indicated in Figure 9 shows that bamboo grains are present in these 3 μm wide lines.

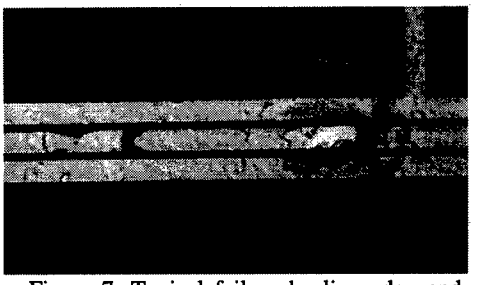


Figure 7. Typical failure by line edge and extended voids in a specimen stressed at 240°C and 3 MA/cm$^2$.

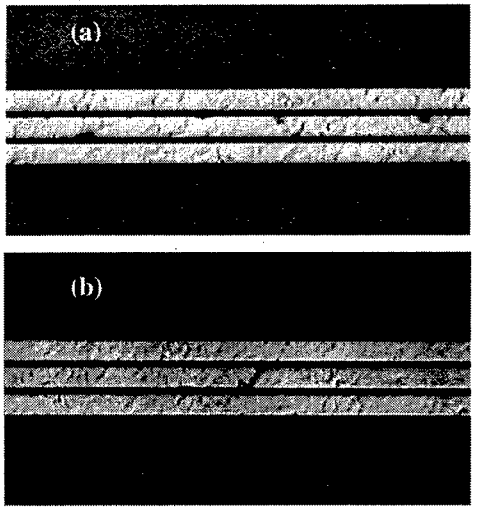



Figure 8. Randomly distributed (a) extended and (b) slit like voids in specimens stressed at 240°C and 2 MA/cm$^2$ and 175°C and 3 MA/cm$^2$.

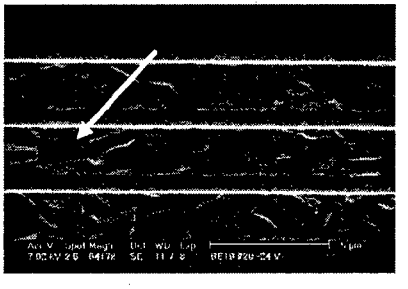


Figure 9. SEM micrograph indicating a bamboo grain in a 3 μm wide line.

Using the results of the failure analysis, the available life-time data was divided into two different populations of 3 μm wide specimens, and the fitting procedure was re-applied to the the weak population only.

Figure 10 shows the results of this analysis. The results were: $E_a = 0.73$ eV, $n = 2.37$, $\sigma = 0.55$, $MTF_0(T_0=125°C, j_0=0.5\ MA/cm^2) =$ 1,088,161 h. The extracted activation energy, very similar to that of the 5 μm wide lines, confirms the validity of our analysis based on two populations. The value of $n = 2.37$ indicates the possible effect of thermal gradients for 3 μm wide lines also, but with a less significant influence than in the wider lines.

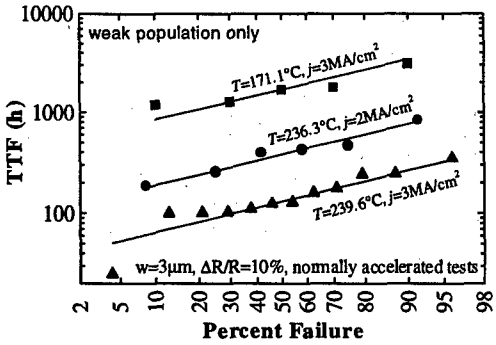



Figure 10. Results of normally accelerated tests on 3 μm wide specimens using only the data of the weak population.

## 4. Conclusions

A comparison between results from normally and highly accelerated EM tests was performed on 5 μm wide samples. Normally accelerated tests were also performed on 3 μm wide specimens. In the case of 5 μm wide lines it was found that the use of Black's equation gives different extrapolated results depending on the range of stress conditions considered. Similar results were obtained with 3 μm wide specimens, provided that a strong population (maybe featuring bamboo sections) was not considered in the analysis. Values of the current density exponent $n$ between 2.4 and 3.3 were found depending on the stress conditions. On the other hand, theoretical treatments have demonstrated that the correct value of $n$ should be 2. This value is based on the assumption of a uniform temperature and current density along the line. However, when high values of current density are used, thermal gradients at defects or at line ends could induce a change of the parameters of the Black equation. In the case of these measurements, thermal gradients at line ends can be excluded, but the presence of thermal gradients at defects or void clusters has been verified. A possible conclusion is that these thermal gradients could increase the value of the current density acceleration factor $n$. In the case of the normally accelerated tests on 5 μm wide and 3 μm wide lines a self heating of about 40°C and 33°C were measured, respectively, and this could be the cause of the value of $n = 2.6$ (2.4, respectively) which was extracted.

A straightforward consequence is that extrapolations made with the Black equation using a current density acceleration factor $n$ which is very different from 2 hardly make sense, because they are dependent on the range of stress parameters adopted [9]. Moreover, the possibility of an incorrect temperature evaluation can be excluded (this is also known to influence the value of $n$ [1]).

The best advice is therefore to use Black's equation only with normal stress conditions (temperatures in the range 150-240°C and current densities between 1 and 2 $MA/cm^2$) and in the case where the extracted value for $n$ is about 2.

Long-term, package-level EM measurements are still needed as a benchmark for highly accelerated measurements. Parameters extracted from fast EM methods which use extremely accelerated stress conditions cannot be safely used for life-time extrapolations to operating conditions.

## 5. Acknowledgments

The authors would like to thank the U.E. SMT Project PROPHECY for financial support and for supplying specimens, Mrs. S. Franceschini, Mr. N. Kelaidis, Mr. A. Sardo, Mr. C. Vezzani, Mr. M. Porcari and Dr. R. Gonella for their contribution to the experiments of this project.

## References

[1] A. Scorzoni et al., Materials Science Reports, 7 (1991) 143.

[2] B. J. Root and T. Turner, Proc. IEEE International Reliability Physics Symposium, 1985, 100.

[3] R. E. Jones, Jr. and Larry D. Smith, J. Appl. Phys., 61 (1987) 4670.

[4] D. G. Pierce, E. S. Snyder, S. E. Swanson and L. W. Irwin, Proc. IEEE International Reliability Physics Symposium (1994) 198.

[5] F. Giroux, C. Gounelle, P. Mortini, G. Ghibaudo, Proc. IEEE Int. Conf of Microelectronics Test Structures, (1995) 229.

[6] "Standard Test Method for Estimating Electromigration Median Time-To-Failure and Sigma of Integrated

Circuit Metallizations", F 1260-89, American Society for Testing and Materials.

[7] Electromigration test procedure proposed in the framework of the PROPHECY SMT Project.

[8] M. Shatzkes and J. R. Lloyd, J. Appl. Phys., 59 (1986) 3890.

[9] J.R.Lloyd, ESREF '97 tutorial notes.

PERGAMON

Microelectronics Reliability 38 (1998) 1029–1034

MICROELECTRONICS RELIABILITY

# Electromigration failure modes in damascene copper interconnects

L. Arnaud[a], R. Gonella[b], G. Tartavel[a], J. Torrès[c], C. Gounelle[b], Y. Gobil[a], Y. Morand[b]

[a] LETI CEA-G, 17 rue des Martyrs, 38054 Grenoble cedex, France
[b] ST Microelectronics, 850 rue Monnet,38926 Crolles, France
[c] CNET France Telecom, chemin du vieux chêne, BP 98, 38243 Meylan, France

**Abstract**

Electromigration experiments were performed on passivated damascene copper interconnects with 1 μm linewidth. A wide range of activation energy values were obtained depending upon barrier layer (Ti or TiN), Cu deposition technique (PVD or CVD process) and grain size. An activation energy of 1.1 eV was measured in PVD-Cu layers leading to significant improvement over AlCu technology : lifetime at 140°C was about 2 orders of magnitude longer. Furthermore, SEM pictures after line failure emphasized interface diffusion mechanisms which occurred in these structures for both Cu CVD and PVD deposition processes.
© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Copper, as a metallization material for future integrated circuits has been actively investigated [1] [2] and takes advantage of its low bulk electrical resistivity, higher melting temperature and mechanical strength. However its reliability performance in comparison to Al alloys remains an issue [3]. Although the bulk and grain boundary activation energy values, are respectively 2.3 and 1.4 eV most of the experimental data obtained from electromigration tests on conventionally etched Cu structures ranges from 0.5 eV to 1.3 eV[4-9]. The highest activation energy value reported was obtained with Cu - PVD deposition process [9]. This process easily provides <111> grain orientation and large grain size. Low activation energies (Ea ~ 0.5eV) obtained in experiments conducted in UHV environment [6] were explained with a mechanism of surface diffusion.

To our knowledge, only few results were reported for electromigration analysis in damascene Cu interconnects [10-11]. Moreover this fabrication process leads to Cu orientation and grain size which depend on the line aspect ratio [12].

In the following study we propose an investigation of electromigration failure modes in damascene Cu interconnects. Median time to failure data and calculated activation energies of passivated narrow Cu lines are compared versus deposition process

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00122-X

(CVD or PVD) and barrier layer composition (Ti or TiN). The results are discussed versus SEM observations of failures and grain sizes. In addition, a comparison of extrapolated lifetimes with data of AlCu lines is proposed.

## 2. Sample preparation

### *2.1. Material deposition*

Damascene Cu lines were fabricated by deposition of Cu into trenches patterned in TEOS oxide. Barrier layers should be added between $SiO_2$ and Cu : it should simultaneously protect the silicon substrate against Cu diffusion, and promote enhanced adhesion at the dielectric / copper interface.

Two barrier layers were studied in this set of experiments : a PVD-Ti barrier (sample#1), which provides good adhesion to both adjacent materials but has generally poor step coverage compared to CVD deposition and a CVD-TiN barrier (samples #2 and #3) which is considered as a better candidate for integration issues [13]. To get a better barrier efficiency, a multistep deposition process followed by a plasma treatment was used.

CVD-Cu deposition used a Cu(tmvs)(hfac) precursor and was performed at 190°C.
For samples #1 and #2 (see Table 1), an 850 nm Cu layer was deposited. Thermal annealing at 400°C, 10 mn was then performed in order to improve adhesion between copper and barrier layer and samples were CMP polished to the final thickness listed in table 2.
For sample#3, before Cu-PVD, a 90 nm seed layer was deposited with the previous CVD process. The PVD deposition was performed at 390°C to allow Cu reflow and obtain a good pattern filling.
The samples were then passivated with 1 μm $SiO_2$ before pad opening. A final annealing at 425 °C for 30 minutes was performed corresponding to the standard annealing step performed after integrated circuit passivation.

### *2.2. Samples characteristics*

The film texture was characterized on passivated blanket films after annealing, with X Ray diffraction using the Bragg method. Figure 1 shows a typical θ–2θ scan obtained with PVD-Cu deposited on CVD-TiN.

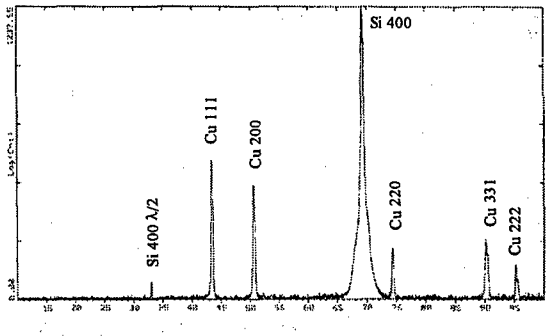



Fig. 1. θ - 2θ scan of PVD Cu deposited on CVD TiN

{111}, {100}, {110}, {311} are diffracted planes of this sample. Considering that the ratio of diffracted intensities are, respectively, for randomly oriented powdered Cu, $I_{200}/I_{111}$ = 0.46, $I_{220}/I_{111}$ =0.20 and $I_{311}/I_{111}$ =0.17, the PVD-Cu sample showed a random orientation.
In the case of CVD-Cu deposited either on PVD-Ti or CVD-TiN barriers, the texture scans θ - 2θ also show many diffracted planes but with $I_{200}/I_{111} \sim 1$. Thus a light (200) texture can be considered. To analyze this texture, the width (2ω) of the (200) peak was measured at a fixed 2θ. A broad (200) peak was obtained in both samples with 2ω = 18°. So we can consider that a poor texture was obtained for this material.
In damascene lines, one can expect that texture will be modified for narrow lines when the line width is smaller than the line thickness [12]. With pole figures of X-Ray diffraction, it is possible to address this issue. Recent experiments performed on passivated CVD damascene copper lines, have shown that the (200) texture obtained in blanket films still remained in lines when the line width is greater than the line thickness.

The grain size of large areas of Cu (pads) and damascene 1 μm lines were revealed by FIB after CMP and final annealing (see Fig.2). The medium grain size and the dispersion were calculated for each sample assuming circular grain shape and lognormal distribution of grain sizes (see Fig.3). One can notice that identical values were obtained for large areas and 1 μm line.
Both CVD-Cu samples show the same medium grain size but the PVD-Cu provides larger grain size due to higher temperature deposition (see Table 1).
Furthermore, FIB imaging of damascene line cross-sections were also conducted for both grain sizes. As shown in Fig. 4, in some places, the damascene

trench is filled with a single grain while in others tilted grain boundaries can be seen providing sometimes small grains located at the line edge.

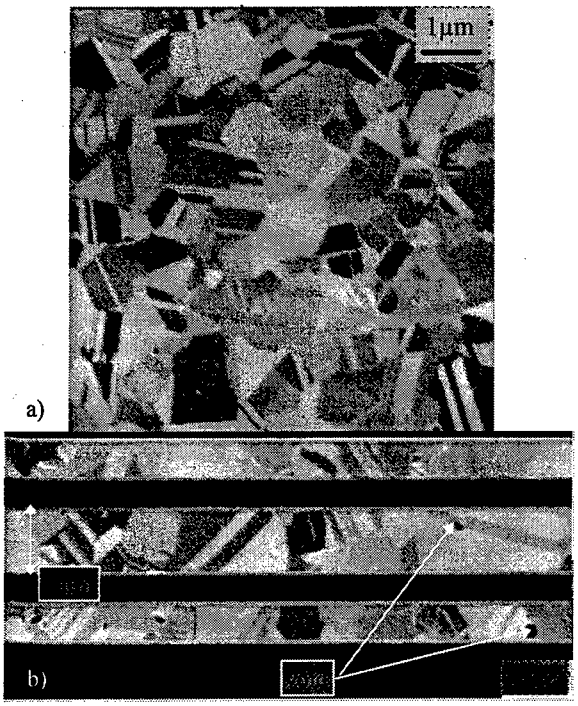



Fig. 2. FIB images of CVD-Cu deposited on CVD-TiN taken a) in a pad area, b)in the damascene 1µm line.

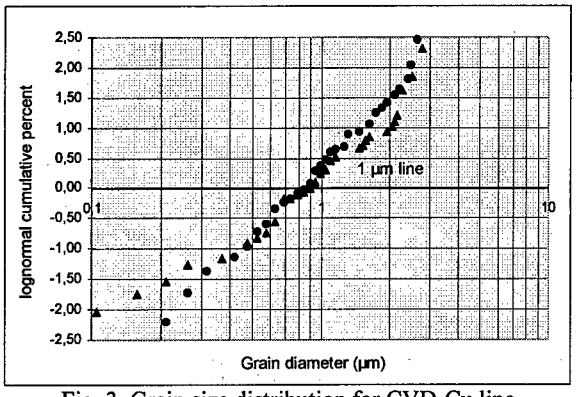



Fig. 3. Grain size distribution for CVD-Cu line and pad (•)

Table 1
Sample description

| Layer characteristics | sample #1 | sample #2 | sample #3 |
|---|---|---|---|
| Barrier layer | PVD Ti | CVD TiN | CVD TiN |
| Cu deposition process | CVD | CVD | PVD |
| Cu texture | poor 200 | poor 200 | random |
| Cu medium grain size | 0.8 µm | 0.8 µm | 1.4 µm |
| dispersion σ | 0.7 | 0.6 | 0.5 |

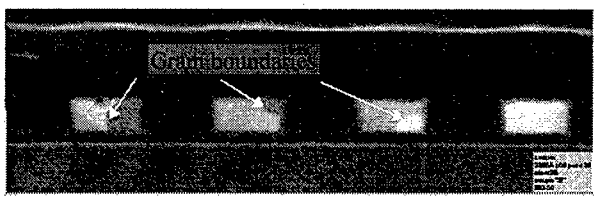



Fig. 4. FIB image of cross-section of 1 µm damascene Cu lines for sample# 2.

## 3. Electromigration experiments

Electromigration test patterns consisted of single level stripes with 4 connecting pads for resistance monitoring. The line length is 3000µm and the connection of the test line to current pads was made with an original « Babel Tower » design, (see Fig 5). Table 2 summarizes the actual geometry of the 3 samples tested.

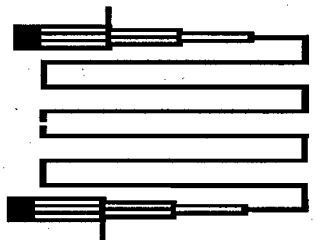


Fig. 5. Electromigration structure.

Table 2
Geometry of electromigration damascene Cu lines.

| Layer geometry | sample#1 | sample # 2 | sample #3 |
|---|---|---|---|
| SiN thickness (nm) | 200 | 200 | 200 |
| Ti or TiN thickness (nm) | 40 | 40 | 40 |
| Cu thickness (nm) | 360 | 425 | 380 |
| $SiO_2$ thickness (µm) | 1 | 1 | 1 |
| Actual line width (µm) | 0.95 | 0.85 | 0.85 |

Highly accelerated tests were performed using the Constant Acceleration Factor Electromigration technique. Typically, these tests are performed at wafer level and at room temperature so that only Joule heating increases the line temperature. In the case of our tests the chuck temperature was maintained at 120°C and a target temperature of about 300°C has been chosen leading to current densities ranging from 27 to 33 MA/cm$^2$. Failure

criterion was 2 % of resistance increase and at this level the stress was stopped and MTF was reported.

Packaged samples were placed in ovens, respectively at 200°C and 250°C and submitted to current densities of 6 and 8 MA/cm$^2$. These rather high current densities (usually Al alloy qualification is made below 3 MA/cm$^2$) were chosen to induce failures within 1000 hours. Similar values were also used by other authors [8]. The increase of line average temperature due to Joule heating was also measured (see Table 3). Due to the low resistivity of Cu, the applied current densities provide less Joule heating than in usual Aluminum reliability testing. Thus, temperature gradients at line extremities were not suspected to induce line failures. Later microscopic observations have confirmed this statement.

Table 3
Average Joule heating for electromigration tests with packaged samples

| Stress conditions To(C)/J(MA/cm2) | sample#1 (°C) | sample#2 (°C) | sample#3 (°C) |
|---|---|---|---|
| 200 / 6 | 12 | 19 | 15 |
| 200 / 8 | 26 | 29 | 27 |
| 250 / 6 | 17 | 20 | 20 |
| 250 / 8 | 31 | 35 | 34 |

## 4. Electromigration results

Figures 6 and 7 show typical graphs of damascene line resistance variation versus electromigration time. Resistance increase as a function of stress time occurred from the beginning of the test for each sample.

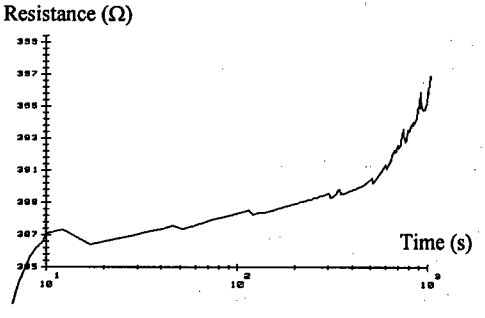



Fig. 6. Resistance versus stress time measured at wafer level for CVD Cu line deposited on CVD TiN.

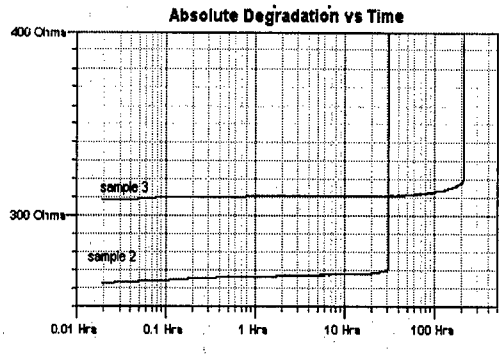



Fig. 7. Resistance versus stress time for packaged Cu lines stressed at 250°C and 6 MA/cm2.

This behavior is generally observed in pure Cu materials while the addition of a small amount of an impurity (Sn) decreases the rate of resistance change indicating that Cu drift velocity is slowed down [14].

Figures 8 and 9 show the time to failure (TF) distributions plotted in a lognormal scale, respectively for highly accelerated tests and packaged sample tests. Median Time to Failure (MTF) and standard deviation ($\sigma$) deduced from the lognormal plots were reported in Tables 4 and 5. For both tests there are significant differences between samples. For given test conditions, Cu-PVD lines exhibit a longer lifetime than CVD lines.

Table 4
Test conditions and electromigration results at wafer level

| **Stress conditions (C)/ (MA/cm2)** | | **sample#1** | **sample#2** | **sample#3** |
|---|---|---|---|---|
| ~ 300 / 30 | MTF(s) | 729 | 780 | 2228 |
| | $\sigma$ | *0.518* | *0.162* | *0.206* |

Table 5
Test conditions and electromigration results for packaged samples

| **Stress conditions (C)/ (MA/cm2)** | | **sample#1** | **sample#2** | **sample#3** |
|---|---|---|---|---|
| ~ 200 / 6 | MTF(h) | 362 | 56 | - |
| | $\sigma$ | *0.25* | *0.38* | |
| ~ 200 / 8 | MTF(h) | 135 | 24 | 790 |
| | $\sigma$ | *0.16* | *0.25* | *0.25* |
| ~ 250 / 6 | MTF(h) | 74 | 43 | 191 |
| | $\sigma$ | *0.17* | *0.21* | *0.10* |
| ~ 250 / 8 | MTF(h) | 24 | 12 | 50 |
| | $\sigma$ | *0.18* | *0.17* | *0.09* |

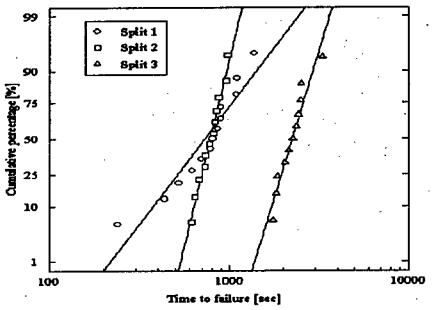



Fig. 8. TF distributions for wafer level tests at 300°C

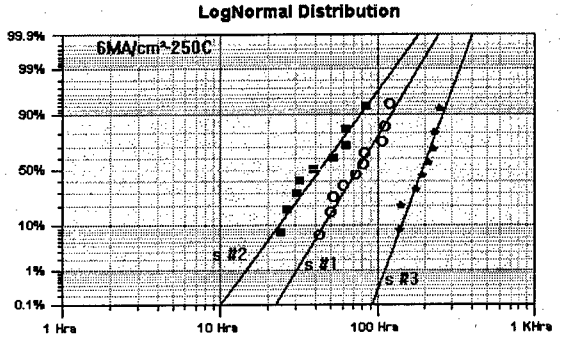



Fig. 9. : TF distributions for packaged samples stressed at 250°C and 6 MA/cm$^2$

From the previous data the parameters of the Black equation (see Eq.1) were computed.

$$MTF = AJ^{-n} \exp\frac{Ea}{kT} \quad (1)$$

We used a global approach described in [15] where the TF of each sample is considered and related to its test conditions by an equation derived from Eq .1. If the TFs are lognormally distributed and tests are completed, the regression gives one set of parameters Ea and n with narrow and realistic confidence intervals named *CI95%* in table 6.

Table 6
Black parameters

| **Black parameters** | **sample#1** | **sample#2** | **sample#3** |
|---|---|---|---|
| Ea (eV) | 0.69 | 0.28 | 1.06 |
| *CI 95%* | *[0.61-0.77]* | *[0.10 -0.40]* | *[0.96-1.16]* |
| n | 2.4 | 2.8 | 2.3 |

Activation energies were ranging from 0.3eV to 1.1eV. Ea for sample#1 (0.7 eV) is a common value of CVD deposition process [10]. The very low Ea for sample#2 (0.28 eV) lets us suppose that surface diffusion occurred in this sample. It also provided the higher n value (2.8). The two other samples had a n factor close to 2, as usually obtained in electromigration experiments when atomic flux divergence is induced by a combination of thermal effect and microstructural effect. n=2 was already measured by some authors in Cu lines etched with conventional process [10].

With the data of Table 5 and 6, we extrapolate the lifetimes of each sample at early failure TF(1%) to the following operating conditions 140°C and 0.2 MA/cm$^2$. Lifetimes of samples #1 and #2 will be respectively 1 000 years and 140 years. Lifetime of PVD-Cu will be as high as 30 000 years. For comparison, the lifetime of 1µm bamboo AlCu interconnects with (111) texture will be 200 years. So PVD-Cu provided longer lifetime, 2 orders of magnitude, than AlCu.

## 5. Failure mode analysis

SEM observations of damascene lines were performed after top $SiO_2$ was removed. Many voids were formed either along line edges (see Fig. 10-a) or with a wedge shape (see Fig.10-b). It has been previously suggested with an in-situ observation [16] that electromigration voiding in PVD Cu conventionally etched was similar to that previously observed in Al : voids nucleated at the line edge, grew laterally, and finally stretched out.

In some cases, as shown in Fig.11, voids do not extend through the whole thickness of the line, thus suggesting that diffusion at the Cu/$SiO_2$ top interface also occurred in this sample as previously proposed for Al electromigration in bamboo structure[17].

Diffusion along the vertical line edges may induce the typical Cu hillocks shown in Fig. 12a. The accumulation of Fig 12-b is a typical defect already encountered in conventional etched AlCu lines where diffusion occurred at the interface between metal and ARC layer [18].

In samples with the low activation energy value, it is suspected that a bad adhesion between Cu and TiN barrier is responsible for an easy Cu diffusion path. Moreover, small voids were seen in FIB images, (see

Fig. 2) located at grain boundary intersections, as previously reported by Keller et al [19]. These voids could also lead to surface diffusion of Cu.
On the contrary, the samples deposited with PVD deposition benefit from the large grain size which provides, besides high activation energy values a low Cu diffusivity.

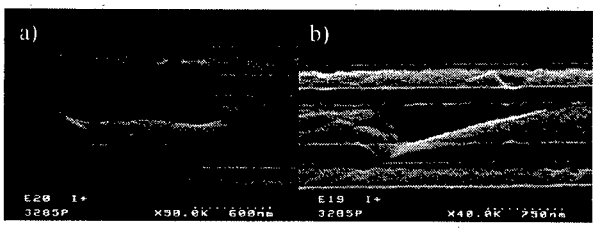



Fig. 10. SEM images of voids in Cu damascene lines of sample#2, after electromigration at T = 250°C, J = 6 MA/cm$^2$ a) void along line edge, b) wedge shape void.

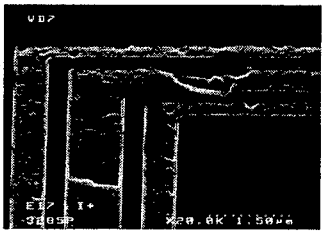


Fig. 11. SEM images of Cu damascene lines of sample#1, after electromigration at T = 250°C, J = 6 MA/cm$^2$

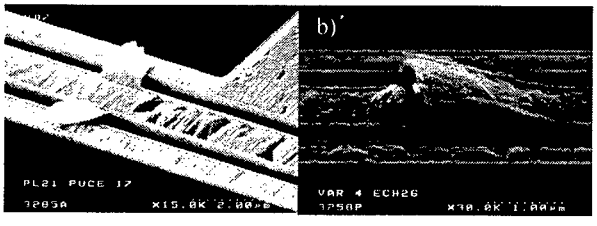



Fig. 12. SEM images of hillocks in Cu damascene lines of sample#3,after electromigration a)T = 243°C, J = 29 MA/cm$^2$ b)T = 250°C, J = 6 MA/cm$^2$

## 6. Conclusion

A detailed investigation of electromigration phenomenon in damascene copper interconnects was made. Median time to failure tests provided lower activation energy values for CVD-Cu deposition process (Ea < 0.7 eV) than for PVD-Cu (Ea = 1.1 eV) although a random Cu orientation was obtained for both processes. In PVD samples the lowest Cu diffusion rate induced extrapolated lifetimes two orders of magnitude longer than AlCu bamboo lines. In the present damascene structures, in addition to the expected grain boundary diffusion, a significant amount of interface diffusion also occurred for both deposition processes.

## Aknowledgments

The authors would like to thank G. Rolland for X Ray experiments, F. Martin for SEM observation, D. Mariolle for FIB imaging.
This work has been carried out within the GRESSI consortium between CEA-LETI and France Telecom CNET and was performed within the framework of the European project DAMASCENE.

## References

[1] D. Edelstein et al, *Proc of IEEE-IEDM Conf (1997), 773.*
[2] J. Torres, *Applied Surface Science 91 (1995), 112-123.*
[3] J. R. Lloyd, J.J. Clement, *Thin Solid Films 262 (1995) 135-141.*
[4] Y. Igarashi, T. Yamanobe, T. Ito, *Jpn. J. Appl. Phys. Vol 34 (1995) 1012-1015.*
[5] C. K. Hu, B. Luther, F.B. Kaufman, J. Hummel, C. Uzoh, D.J. Pearson, *Thin Solid Film 262 (1995), 84-92.*
[6] R.W.Vook, B.S. Ho, *Mat. Res. Soc. Symp. ,vol 428, (1996), 31.*
[7] F. Braud, G. Tartavel, J. Palleau, J. Torrès, A. Persico, G. Reimbold, *Proc of ESREF Conf (1995), 235*
[8] C. Ryu, A.L.S. Loke, T. Nogami, S. Wong, *Proc of IEEE IRPS (1997), 201.*
[9] S. Shingubara, *Proc. of Electrochemical Soc. Meeting (1996),vol 96-2, 262.*
[10] G. Minamihaba et al, *Jpn. J. Appl. Phys. Vol 35 (1996) 1107.*
[11] G. Bai et al, *Proc. of Symp. on VLSI Technology, (1996), 48.*
[12] K. Abe, Y. Harada, H. Onoda, *Proc. of IEEE-IRPS, (1998), 342.*
[13] M. Moussavi, Y. Gobil, et al, *Proc. of IEEE-IITC Conf. (1998), p295.*
[14] C.K. Hu, K.L. Lee, D. Gupta, P. Blauner, *Mat. Res. Soc. Symp. Proc, vol 428, (1996), 43.*
[15] P. Waltz, L.Arnaud, G. Lormand, G. Tartavel, *to be publised in Microelectron. Reliab.*
[16] S-H. Lee, J.C. Bravman, P. Flinn, L. Arnaud, *Mat.Res. Soc. Symp. Proc., vol 473 (1997), 235-239*
[17] R.A. Augur, F. Van den Elshout, R.A.M. Wolters, *AIP. Conf Proc., Vol 373, (1996), 279-292.*
[18] T. Kouno, M. Hosaka, H. Niwa, M. Yamada, *J. Vac. Sci, Technol. Vol B15, (1997), 1800.*
[19] R.R Keller, J.A. Nucci, D.P. Field, *Journal of Electronic Material, vol 26, (1997), 996.*

PERGAMON

Microelectronics Reliability 38 (1998) 1035–1040

MICROELECTRONICS RELIABILITY

# The dependence of stress induced voiding on line width studied by conventional and high resolution resistance measurements

A. Witvrouw[a], K. Maex[a,b], W. De Ceuninck[c], G. Lekens[c], J. D'Haen[c] and L. De Schepper[c]

[a] *IMEC, Kapeldreef 75, 3001 Leuven, Belgium.*
[b] *INSYS, K.U. Leuven, Belgium.*
[c] *IMO, Wetenschapspark 1, 3590 Diepenbeek, Belgium.*

**Abstract**

The degradation due to stress induced voiding of nitride passivated Al-1 wt.% Si and Ti/TiN/ Al-1wt.% Si-0.5 wt. % Cu/Ti/TiN interconnects with widths ranging between 0.4 and 1.2 μm was studied by in-situ conventional and high resolution resistance measurements (HRRM) during storage at temperatures between 168 and 240 °C. The conventional measurements on Al-Si lines, which lasted more than one year, clearly showed that the interconnect lifetime decreases with decreasing line width. With HRRM the degradation due to stress induced voiding can be detected much sooner and with much more detail. From the HRRM it is clear that the resistance changes during storage often happen in jumps and that the degradation has a rather complex alloy, line width and temperature dependence. Both for 0.4 and 0.6 μm wide Al-Si lines more degradation occurred for storage at 175 °C compared to storage at 200 °C. For the Al-Si-Cu stacks the degradation of 0.4 μm wide lines was worse for storage at 240 °C compared to storage at 200 °C, but the opposite was true for the 0.6 μm wide lines. © 1998 Elsevier Science Ltd. All rights reserved.

## Introduction

Stress induced voiding in an Al-based conductor is caused by the migration of the Al-ions under influence of the high tensile stresses which can exist in the interconnect after processing. This voiding can eventually lead to an electrical open and thus failure of the circuit. The tensile stresses are caused by the difference in thermal expansion between the Al and the Si-substrate on one hand and the Al and the dielectric isolation material or passivation layer on the other hand.

This tensile stress increases strongly with decreasing dimensions and therefore stress induced

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00119-X

voiding as well. For sub micron lines models [1] even predict an increase in lifetime with the fifth power of the line width and the third power of the line thickness. This means that stress induced voiding should be of a growing concern when technology progresses and miniaturisation continues.

In this work the lifetime of nitride-passivated Al-Si and Al-Si-Cu lines of different line width was electrically evaluated by using both conventional and high resolution resistance (HRRM) measurements. Initially Al-Si was chosen over Al-(Si)-Cu to reduce the measurement times. However, as the high sensitivity of HRRM allows for shorter measurement times, stress induced voiding in nitride passivated Ti/TiN/Al-Si-Cu/Ti/TiN lines of different line width was studied by the use of HRRM as well.

## Experimental procedure

400 nm of Al-1 wt.%Si and a stack of 20 nm Ti/ 80 nm TiN/ 500 nm Al- 1 wt.% Si-0.5 wt.% Cu/ 20 nm Ti/ 60 nm TiN were sputter deposited on thermally oxidized Si-wafers. 2 mm long 4-point resistors of different line width were defined with i-line lithography and conventional RIE. The line width of the lines used in this work was 0.4, 0.6, 0.8 and 1.2 μm for Al-Si and 0.4 and 0.6 μm for Al-Si-Cu. After line definition, a 2 μm and 1 μm, respectively, thick Si-nitride passivation was deposited at 390 °C on top of the Al-Si and Al-Si-Cu lines, respectively. A final sinter step for 20 min. at 420 °C in 10%H2/N2 was given.

The in-situ resistance measurements were performed on packaged samples. For the conventional resistance measurements of Al-Si lines 96 DUTs (Devices Under Test) , 12 of each line width at each of the two test temperatures, were stored at 168 °C for 12016 h and at 197 °C for 11710 h. For the HRRM on Al-Si 123 DUTs, 31 and 30, respectively, 0.4 μm wide lines and 32 and 30, respectively, 0.6 μm wide lines were stored at 175 and 200 °C, respectively, for 1200 h. For the HRRM on Al-Si-Cu 42 DUTs, 12 and 9 DUTs, respectively, of each line width were stored for 2120h at 200 °C and 240 °C, respectively. All HRRM measurements were performed in a DESTIN electromigration system. The typical resolution of this system is 80 ppm for Al-based interconnects. For the Al-Si lines the local temperature correction algorithm option was implemented, resulting in an even improved resolution of 10 ppm.

## Results and discussion

### *Conventional Resistance Measurements on Al-Si*

Apart from the initial resistance decrease due to precipitation [2], the resistance signal stays rather flat (within the noise) until failure. Failure, when using conventional resistance measurements, is always an electrical open. It is not so surprising that only open circuits can be detected, as even large voids contribute only very little to the resistance increase. For example, a void of 1 μm length and with a width equal to half of the conductor width will only give rise to a resistance increase of 500 ppm or 0.05% in a 2 mm long line. And actually the resistance decrease will be even lower because of the decrease of the resistivity with decreasing tensile stress [3]. Therefore measurement times, when using conventional resistance measurements, have to be very long in order to evaluate failure. Table 1 gives an overview of the # failures per temperature and line width after more than a year (!) of storage. Almost all failures are 0.4 μm lines and no 1.2 μm line failed. It is also interesting to see that the number of failures does not depend much on the chosen test temperature. Failure data of 0.4 μm lines at both temperatures was fitted with lognormal distributions with the same deviation in time to failure ($\sigma$). This resulted in a $\sigma$ of 1.3 and a MTTF of 11000 h at 168°C and 12587 h at 197 °C. This implies a very

low activation energy if a model from literature is used [4,5].

Table 1
Overview of the # of failed (or to censor) devices as a function of test temperature and line width.

| T (°C) | 0.4 μm | 0.6 μm | 0.8 μm | 1.2 μm |
|---|---|---|---|---|
| 168 | 6 (1) / 12 | 0 / 12 | 1(1) /12 | 0 / 12 |
| 197 | 5 / 12 | 1 (1) / 12 | (1) / 12 | 0 / 12 |

*High Resolution Resistance Measurements on Al-Si*

Because of the high resolution a much greater sensitivity towards changes in the resistance is obtained. Even during a test of 1200 h 4 different groups of samples can be distinguished. Group 1 has almost no change in resistance during the test (Fig. 1a). SEM microscopy revealed no or only very small defects in the lines after the test. Group 2 has a very slow continuous monotone increase in resistance for a total of 700 to 2000 ppm change in 1200 h (Fig. 1b). In the SEM several non-critical voids (Fig. 2a and 2b) were seen over the whole length of the lines. The samples from group 3 have an abrupt step in the resistance versus time behavior (Fig. 1c). This step results from the abrupt growth of a large critical void (Fig. 3). The magnitude of the step ranges between 50 and 200 ppm and is correlated to the size of the void. Although these voids do not give rise to a large resistance increase, it is clear that they can have a detrimental effect on the lifetime in a subsequent electromigration test [3]. Finally, the samples from group 4 failed from an electrical open (Fig. 1d). This happened to only 7 samples out of 123. The distribution of the different types of failures for all measured samples is listed in Table 2. The boundary between type 1 and type 2 failures is set at 500 ppm resistance increase. From table 2 it is clear that there is less degradation at 200°C compared to 175 °C as the percentage of type 1 failures is higher at that temperature. However, not much temperature dependence can be seen for the hard failures or type 4 failures, which occur only in 0.4 μm wide lines. This is consistent with the median time to failure values found from the conventional testing. On the other hand, 0.6 μm lines seem to be more susceptible to non-critical damage (type 2 and 3 failures). However, for these wider lines this damage is not converted to a hard failure as quickly.

Table 2
Distribution of the different failure types .

| line width | 0.4 μm | 0.4 μm | 0.6 μm | 0.6 μm |
|---|---|---|---|---|
| test T | 175 °C | 200 °C | 175 °C | 200 °C |
| type 1 | 29 % | 70 % | 38 % | 57 % |
| type 2 | 55 % | 7 % | 50 % | 27 % |
| type 3 | 6 % | 10 % | 12 % | 16 % |
| type 4 | 10 % | 13 % | 0 % | 0 % |

*High Resolution Resistance Measurements on Al-Si-Cu*

From a previous study [3] it is known that the resistance changes due to stress induced voiding in Al-Si-Cu stacks occur in jumps. Each jump indicates the time a void crosses the complete width of the line. For these voids measurable resistance increases are expected as the current is forced through the barrier and anti-reflective coating (ARC). As voiding changes the stress in the remaining Al and as this stress change causes a resistance change, the total resistance increase $\Delta R$ due to a void with length $\Delta l$ can be written as [3]:

$$\Delta R = \Delta l \cdot (R_{(\text{barrier+ARC})/\text{unit length}} - 1.93 \cdot R_{\text{Al/unit length}}) \quad (1)$$

and

$$\Delta R/R = \Delta R/(R_{\text{Al/unit length}} \cdot 2000) = \Delta l \cdot X \quad (2)$$

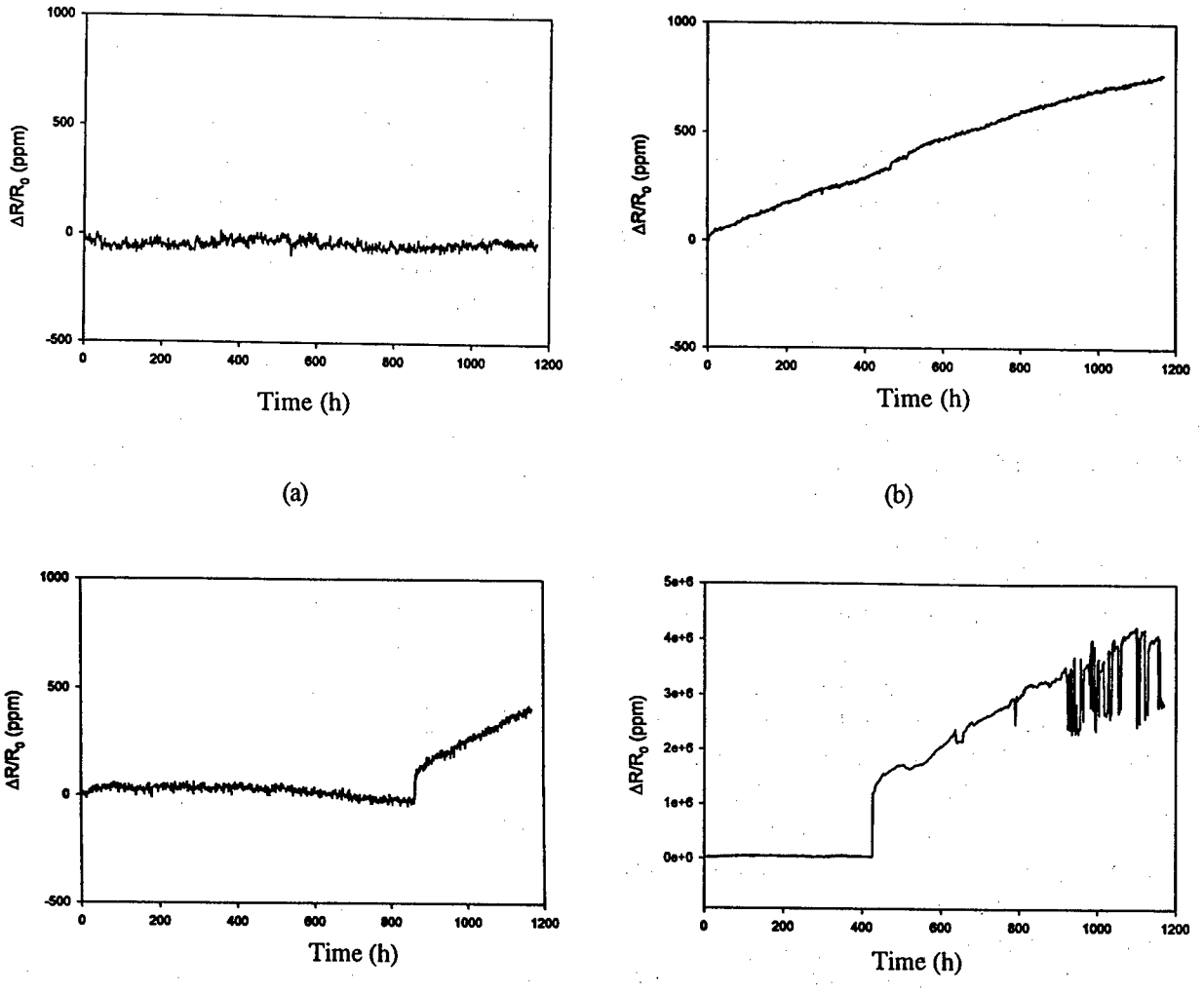
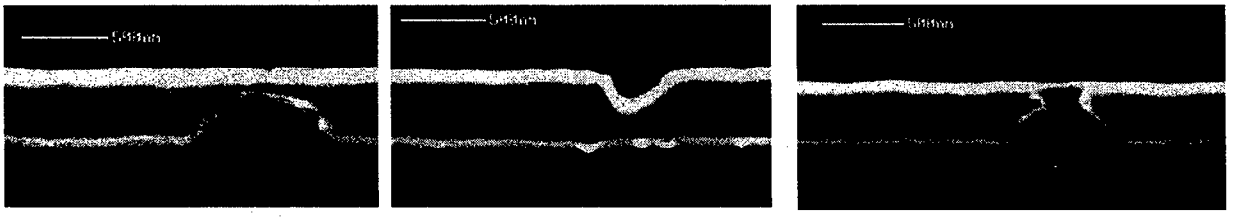



Fig. 1. Four types of failures in Al-Si lines: type 1: no change for an 0.6 μm line stored at 175 °C (a); type 2: monotone increase in resistance in an 0.4 μm line stored at 175 °C (b); type 3: step in resistance behavior in an 0.4 μm line stored at 200 °C (c) and type 4: open contact for an 0.4 μm line stored at 200 °C (d).



(a) (b)

Fig. 2a and 2b. Two examples of non-critical voids found in the 0.4 μm line with the resistance behavior of Fig. 1b.

Fig. 3. Large critical void in the 0.4 μm line with the resistance behavior of Fig. 1c.

with X = 0.018 Ω/μm at 200 °C and X = 0.017 Ω/μm at 240 °C for 476 nm of Al-Si-Cu, a barrier of 20 nm Ti/ 80 nm TiN and an ARC of 30 nm $TiAl_3$/ 12 nm Ti/ 60 nm TiN. Thus a void with a width of 1 μm will result in a 1.8 % or 18000 ppm resistance increase at 200 °C and a 1.7 % or 17000 ppm resistance increase at 240 °C. The results of all measurements at 200 °C and 240 °C are shown in Figs. 4a, 4b, 4c and 4d. From Fig. 4 it is clear that there is quite a lot of spread in the resistance data for a certain test condition. Also here jumps or steps are observed, indicative of newly formed voids, crossing the width and height of the conductor. As new such voids are formed at different times during the test, the resistance data does not show the simple square root of time dependence predicted by models [6].

Comparing the results from both test temperatures, it is seen that an anneal at 240 °C leads to more degradation compared to an anneal at 200 °C for the narrow lines, while the opposite is true for the wider lines. This was also seen by the use of other measurement techniques [7]. The maximum resistance increase (and thus the maximum void length) at 200 °C is actually even higher for 0.6 μm wide lines compared to 0.4 μm wide lines, which might be caused by a higher diffusivity and/or more voids for the wide lines [6]. However, the percentage of lines showing degradation is lower for 0.6 compared to 0.4 μm. This is certainly true at 240 °C where only 2 out of 9 tested 0.6 μm wide lines have some degradation. The resistance of the 7 other DUTs actually continuously decreases until the end of the test. This resistance decrease has also been reported for

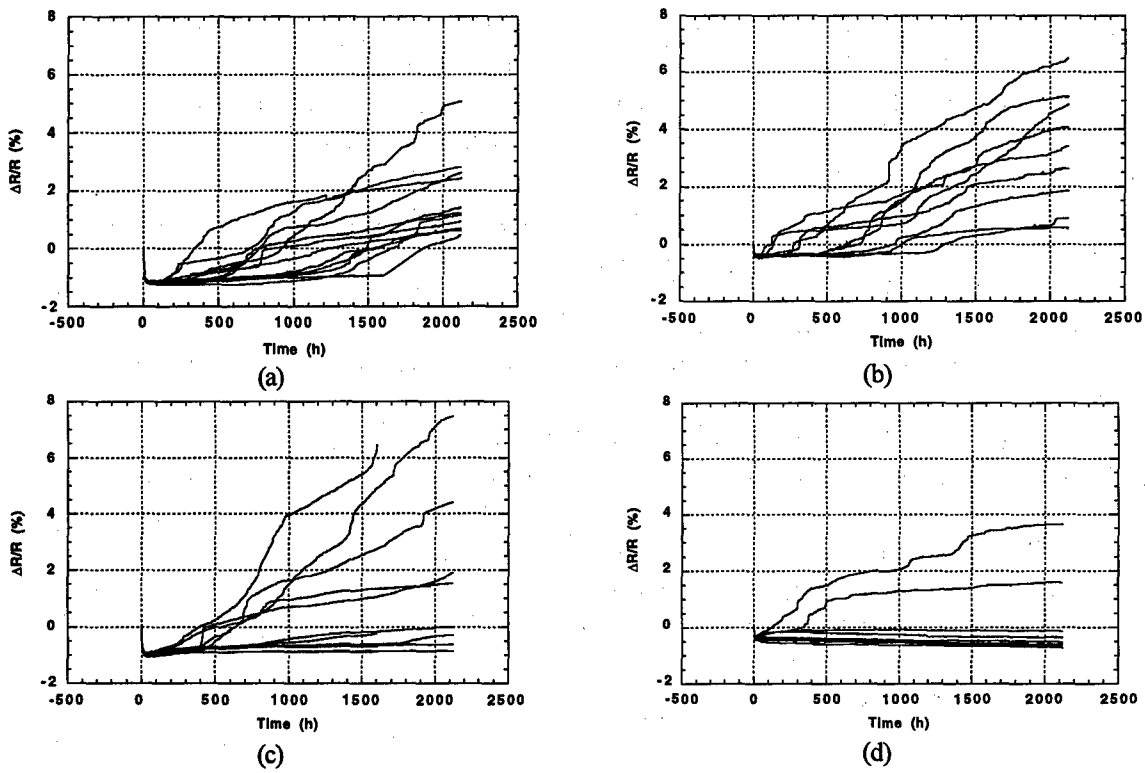



Fig. 4. HRRM of Ti/TiN/Al-Si-Cu/Ti/TiN lines with a line width of 0.4 μm stored at 200 °C (a) and 240 °C (b) and with a line width of 0.6 μm stored at 200 °C (c) and 240 °C (d).
A very low current stress of 100 μA was used for the resistance measurements.

electromigration tests on Al-stacks with a Ti/TiN ARC [8], where it was found that thickening of the tested line occurred due to Al-diffusion from the bonding pad as the $TiAl_3$ layer is considered a fast diffusion path. Perhaps a similar mechanism is operative in the 0.6 μm wide lines stored at 240 °C.

## Conclusion

When failure is defined as an electrical open, it is clear from conventional resistance measurements that the stress induced voiding lifetime of passivated Al-Si lines decreases for decreasing line width. Moreover, testing at 168 or 197 °C resulted in comparable lifetimes.

With HRRM the degradation due to stress induced voiding can be detected much sooner and with much more detail. From the HRRM it is clear that this degradation has a rather complex behavior.

Al-Si samples could be divided into 4 groups based on their HRRM behavior. From this it was clear that the 0.6 μm wide lines had more non-critical damage compared to 0.4 μm wide lines. However, in contrast to the 0.4 μm wide lines, this did not lead to hard failures during the duration of the test Also, with exception of the hard failures, degradation was larger at 175 °C compared to 200 °C. This is in agreement with earlier reports on Al-Si interconnects [4].

HRRM on Al-Si-Cu stacks indicate a different temperature dependence of the degradation due to stress induced voiding for 0.4 and 0.6 μm wide lines. The degradation of the narrow lines is worse at 240 °C compared to 200 °C and the opposite is true for the wider lines. This difference could be due to a difference in diffusion mechanism in 0.4 and 0.6 μm wide lines.

## Acknowledgement

The authors thank the IMEC pilot line for processing, J. Van Laer for technical assistance and Ch. Drijbooms for FIB imaging. The Al-Si work was performed under a contract with Alcatel Bell, Antwerp, Belgium and with support of the Flemish Institute of Scientific and Technological Research. The Al-Si-Cu work was partly funded by the European project ACE.

## References

[1] H. Okabayashi, Proc. ESREF '91, 171 (1991).
[2] W.A. De Ceuninck, V. D'Haeger, J. van Olmen, A. Witvrouw, K. Maex and L. De Schepper, P. De Pauw and A. Pergoot, Microelectronics Reliability, 38 (1), 87-98 (1998).
[3] A. Witvrouw, Ch. Drijbooms, H. Bender and K. Maex, Proc. Advanced Metallization and Interconnect Systems for ULSI Applications in 1997, 279 (1998).
[4] J.W. McPherson and C.F. Dunn, J. Vac. Sci. Tech. B5, 1321 (1987).
[5] A.I. Sauter and W.D. Nix, J. Mater. Res. 7 (5), 1133 (1992).
[6] C.K. Hu, K.P Rodbell, T.D. Sullivan, K.Y. Lee and D.P. Bouldin, IBM J. Res. Develop. 39 (4), 486 (1995).
[7] A. Witvrouw, unpublished results.
[8] T. Kouno, M. Hosaka, H. Niwa and M. Yamada, J. Vac. Sci. Technol. B 15 (5), 1800-1804 (1997).

PERGAMON

Microelectronics Reliability 38 (1998) 1041–1046

MICROELECTRONICS RELIABILITY

# Lateral interface effect on pulsed DC electromigration analysis

**P. Waltz[a], L. Arnaud[a], G. Lormand[b], G. Tartavel[a]**

[a] *CEA LETI - 17 Av. des Martyrs - 38054 GRENOBLE Cedex 09 - France*

[b] *GEMPPM - INSA - UMR CNRS 5510, Bt 303 - 69621 VILLEURBANNE Cedex - France*

**Abstract**

DC and Pulsed DC Electromigration tests at 1Hz and 10kHz have been performed on two single level Al-0.5%Cu metallizations. The Black's parameters have been analysed with a great confidence level using a statistical global approach. The results are in agreement with the Average Current Model at 10kHz and with the On-Time Model at 1Hz considering that thermal effects not only affect the current density exponent n, but also the duty cycle accelerating factor m. The extracted activation energies reflect the same diffusion mechanisms for the two metallizations. Microscopic observations showed huge metal accumulations for each structure and emphasized the influence of the resist stripping stage on the electromigration behaviour of the samples.
© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

DC and Pulsed DC (PDC) electromigration (EM) tests at 1 Hz and 10 kHz have been performed on structures from two slightly different batches. From the generalized Black's equation [1,2,3],

$$MTF = AJ^{-n}r^{-m}\exp(Q/kT) \quad (1)$$

we could extract the current density and duty cycle accelerating factors, n and m, and the activation energy Q in order to validate the models proposed in literature [2,3,4] (On-time model and Average current model). A multilinear regression has been used, after temperature calculation with an original thermal model [5], in order to calculate narrow and realistic confidence intervals of the Black's parameters and of the dispersion σ. Moreover, microstructural characterisation shows that, for one batch, the experimental resistance decrease can be perfectly correlated to the size of huge metal accumulations in the line. For the second batch, the accumulations still exist but the resistance decrease is compensated by an increase due to a voiding effect at the lateral surfaces of the metal line. This different EM behaviour has to be related to a different stripping process.

## 2. Experiments

The tests were performed on single level AlCu NIST structures. The thickness of the layers from top to bottom is given in Table 1. The 800μm long and 3μm wide lines are connected to the pads via a 6μm wide current access line. The grain size of the metal (D=2.8μm) is close to the width of the line and so we can assume it to be in a quasi-bamboo configuration. The difference between the two batches consists in a little thicker underlying oxide layer for structure 1 (See Table 1) and in a further step during stripping of the lines for Structure 2.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00116-4

Table 1
Thickness of the layers for the tested structures

| Layers | Structure 1 | Structure 2 |
|---|---|---|
| $SiO_2$ | 700 nm | 700 nm |
| TiN | 40 nm | 40 nm |
| Ti | 10 nm | 10 nm |
| AlCu | 440 nm | 440 nm |
| TiN | 60 nm | 60 nm |
| Ti | 40 nm | 40 nm |
| $SiO_2$ | 600 nm | 500 nm |
| Si | Substrate | Substrate |

For each stress condition (See Table 2 and Table 3), we tested a batch of 16 samples. We considered the failure of the line after a 40% resistance increase.

## 3. Statistical analysis

The Black's parameters are usually extracted from a succession of linear regressions which lead to describe a same test by as many values of n as different pairs (T ;r), as many values of Q as different pairs (J ;r) and as many values of m as different pairs (T ;J). To that, we add the problem concerning the calculation of confidence intervals. By using the MTF's instead of the time-to-failures, these intervals are unrealistically reduced due to an important decrease in the random feature of the results.

The proposed globalized approach uses the same hypothesis than the usual methods [6] but with a different formulation. The Time to Failure (TF) of each sample is related to its test conditions by the following equation derived from Equation 1 :

$$\ln TF = \ln B - n\ln J - m\ln r + Q/kT_{eff} \qquad (2)$$

If the TFs are lognormally distributed [7], then the sample parameter B follows also a lognormal distribution with mean A (See Eq. 1) and standard deviation σ. If the tests are complete (all the samples failed), all the conditions are verified for a multilinear regression analysis. This method is used over all the TFs, all the current densities, all the duty cycles and all the effective temperatures using Equation 2. Such a regression gives one set of parameters A, n, m, Q and σ, and it's statistical properties [8] allow to calculate narrow and realistic confidence intervals because of the large population considered. The confidence interval of σ is extremely important, especially to extrapolate lifetimes to low failure rates.

## 4. Results and analysis

From DC tests and tests at 10kHz for Structure 1 and 1Hz for Structure 2, we extracted, on one hand, the MTFs (Median Time to Failure) and the standard deviations from the plots Log(TF) versus Cumulative failure (See Table 2 and Table 3), and, on the other hand, the usual Black's parameters using the described statistical approach. The results are summarized in Table 4 and Table 5.

Table 2
Test conditions and results for *Structure 1* (10 kHz)

| Conditions | $T_{oven}$= | 200°C | 250°C |
|---|---|---|---|
| J=2 MA/cm² | MTF= | ***881.5 h*** | ***115.7 h*** |
| DC | σ= | *0.64* | *0.24* |
| J=4 MA/cm² | MTF= | ***76.7 h*** | ***19.3 h*** |
| DC | σ= | *1.09* | *0.61* |
| J=6 MA/cm² | MTF= | ***6.1 h*** | ***2.6 h*** |
| DC | σ= | *0.61* | *0.67* |
| J=4 MA/cm² | MTF= | ***477.8 h*** | ***80.5 h*** |
| r=50% | σ= | *0.13* | *0.19* |
| J=4 MA/cm² | MTF= | ***170.7 h*** | ***31.3 h*** |
| r=80% | σ= | *0.85* | *0.48* |
| J=6 MA/cm² | MTF= | ***607.5 h*** | ***182.1 h*** |
| r=20% | σ= | *0.31* | *1.14* |
| J=6 MA/cm² | MTF= | ***47.7 h*** | ***16.1 h*** |
| r=50% | σ= | *1.07* | *0.89* |
| J=6 MA/cm² | MTF= | ***10.0 h*** | ***3.2 h*** |
| r=80% | σ= | *1.16* | *0.59* |

Table 3
Test conditions and results for *Structure 2* (1 Hz)

| Conditions | $T_{oven}$= | 200°C | 250°C |
|---|---|---|---|
| J=3 MA/cm² | MTF= | ***478.8 h*** | ***81.0 h*** |
| DC | σ= | *0.87* | *0.12* |
| J=5 MA/cm² | MTF= | ***15.1 h*** | ***5.2 h*** |
| DC | σ= | *0.41* | *0.34* |
| J=4 MA/cm² | MTF= | ***720.6 h*** | ***79.9 h*** |
| r=50% | σ= | *0.19* | *0.25* |
| J=4 MA/cm² | MTF= | ***271.7 h*** | ***30.2 h*** |
| r=80% | σ= | *0.76* | *0.41* |
| J=6 MA/cm² | MTF= | ***16.7 h*** | ***5.0 h*** |
| r=50% | σ= | *0.72* | *0.48* |
| J=6 MA/cm² | MTF= | ***7.2 h*** | ***2.0 h*** |
| r=80% | σ= | *0.63* | *0.41* |

The parameters given in the pulsed case are extracted using all the results, including DC results with a singular value of duty cycle (r=100%). It is then possible to extrapolate each TF to the same stress conditions and to extract a single value of MTF and σ from a global test including all the 16 tests for Structure 1 (See Fig. 1). The confidence interval at 95% for each sample is also represented.

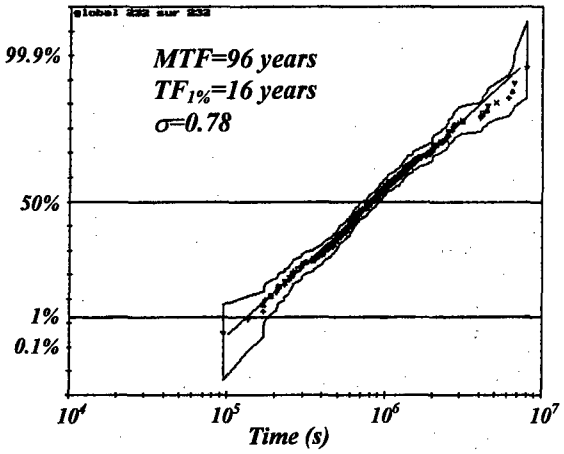



Fig. 1. Extrapolation to T=140°C, J=0.2MA/cm² and r=1

On one hand, to a confidence level of 95%, we found, for each structure, $Q_{dc}$=$Q_{pdc}$ and $n_{dc}$=$n_{pdc}$. Consequently, we have no argument to consider different diffusion mechanisms in DC and PDC cases. On the other hand, the activation energies for Structure 1 and Structure 2 are very close, especially in DC case, giving evidence of a similar diffusion mechanism for both metallizations.

Table 4
Black's parameters given by the multilinear regression and their confidence intervals (CI) at 95% for ***Structure 1***

| DC Tests | Tests at f=10kHz |
|---|---|
| **$Q_{dc}$ = 0.66 eV** | **$Q_{pdc}$ = 0.69 eV** |
| *CI95% [ 0.52 - 0.80 ]* | *CI95% [ 0.59 - 0.78 ]* |
| **$n_{dc}$ = 2.63** | **$n_{pdc}$ = 2.63** |
| *CI95% [ 2.18 - 3.07 ]* | *CI95% [ 2.27 - 2.99 ]* |
| - | **$m_{pdc}$ = 2.76** |
| - | *CI95% [ 2.54 - 2.98 ]* |
| **σ = 0.73** | **σ = 0.78** |
| *CI95% [ 0.64 - 0.85 ]* | *CI95% [ 0.71 - 0.86 ]* |

Table 5
Black's parameters given by the multilinear regression and their confidence intervals (CI) at 95% for ***Structure 2***

| DC Tests | Tests at f=1Hz |
|---|---|
| **$Q_{dc}$ = 0.68 eV** | **$Q_{pdc}$ = 0.78 eV** |
| *CI95% [ 0.55 - 0.81 ]* | *CI95% [ 0.70 - 0.86 ]* |
| **$n_{dc}$ = 4.61** | **$n_{pdc}$ = 4.87** |
| *CI95% [ 3.99 - 5.24 ]* | *CI95% [ 4.48 - 5.27 ]* |
| - | **$m_{pdc}$ = 2.30** |
| - | *CI95% [ 1.98 - 2.61 ]* |
| **σ = 0.52** | **σ = 0.57** |
| *CI95% [ 0.44 - 0.64 ]* | *CI95% [ 0.52 - 0.64 ]* |

The calculated activation energies and the quasi-bamboo microstructure give rise to a competition between grain boundary diffusion and interface diffusion [9].

For Structure 1 at 10kHz, we found $n_{pdc}$=$m_{pdc}$. This result validates the « Average Current Model » for which the sample is stressed by an efficient current density averaged over the period *p*. Equation 1 can then be written as follows :

$$MTF = AJ_{eff}^{-n} \exp(Q/kT) \qquad (3)$$

In the case of the ACM model, this efficient current density is defined by Equation 4.

$$J_{eff} = \frac{1}{p}\int_{p} J(t)dt = rJ \qquad (4)$$

The obtained value, higher than the commonly used value of 2, appears as a result of the competition between structural induced (n=1) and thermal induced (n=3) failures [10]. English and Kinsbron also found that thermal effects like Joule heating affect the duty cycle accelerating factor m [11].

For Structure 2 at 1Hz, the results are different ($m_{pdc}$<$n_{pdc}$) and tend towards the « On-Time Model ». For this model, also called RMS model, the efficient current density is given by Equation 5 [12].

$$J_{eff} = \left[\frac{1}{p}\int_{p}\left(J(t)\right)^2 dt\right]^{1/2} = \sqrt{r}J \qquad (5)$$

By combining Equation 5 with Equation 3, we found that $m_{pdc}=1/2\ n_{pdc}$. This assumption correlates perfectly with the experimental results (See Table 5). The high values of n and m for Structure 2 will be related later to the microscopic observations.

Furthermore, many SEM observations have been performed after test and removal of the passivation layer. As shown in Figure 2 and Figure 3, we found huge accumulations all along the line for both structures. For Structure 1, the accumulation dimensions have been perfectly correlated to the experimental resistance decrease shown in Figure 4 [13]. For these samples, there is no depletion in line and all the matter comes from the current access line [13,14]. Seeing the shape of the accumulations (See Fig. 2 and Fig. 3) and the access line depletion, the Ti/AlCu interface seems to be the more active diffusion path for both structures as already reported in similar samples [15,16,17].

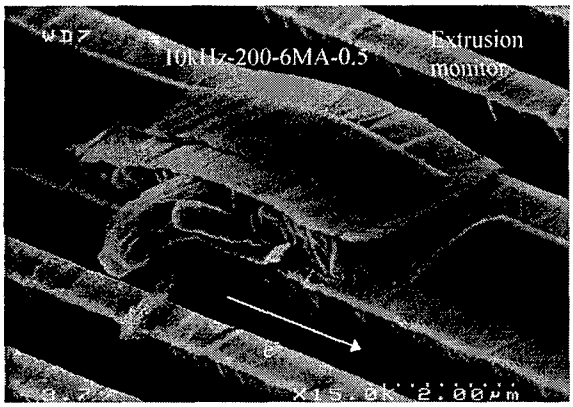


Fig. 2. SEM view of an accumulation - *Structure 1*

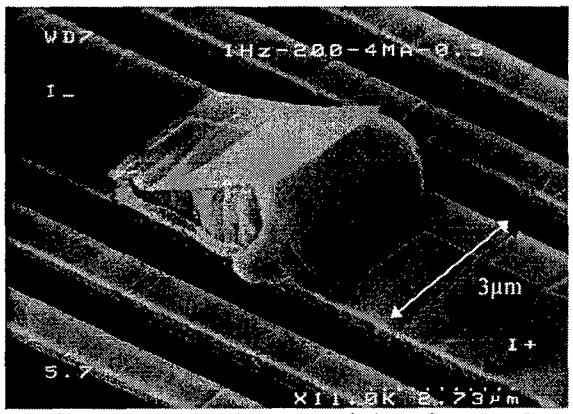


Fig. 3. SEM view of an accumulation - *Structure 2*

The matter travels along this interface as long as it is not blocked by a change in the orientation of the underlying aluminium grain inducing a decrease of the interface diffusion coefficient. This effect results then in local homogeneous grain growth along z-axis (See Fig. 2 and Fig. 3).

As shown in Figure 4, Structure 2 does not show a resistance decrease even if there are accumulations in the line.

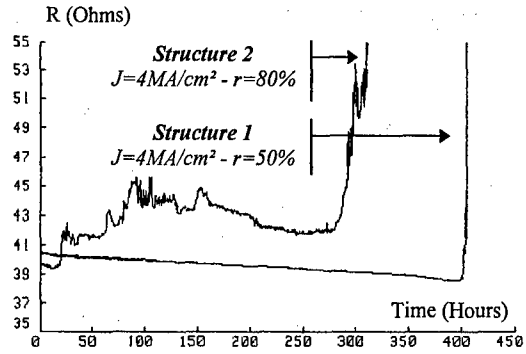


Fig. 4. Resistance vs Stress time - $T_{oven}$=200°C

Figure 5 shows a lot of lateral voids in the line itself and also in the 6µm wide current access line due to entire grains diffusing at the interfaces and accumulating further in the line.

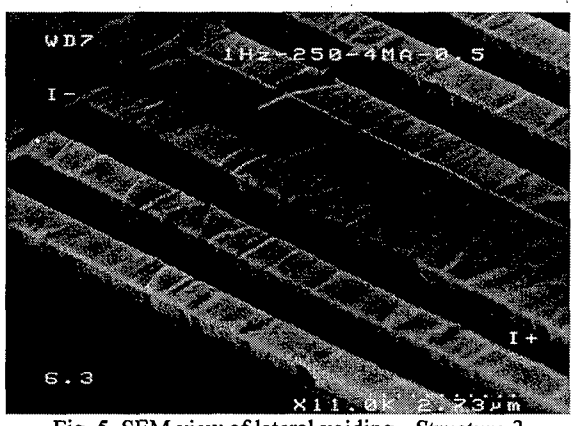


Fig. 5. SEM view of lateral voiding - *Structure 2*

The appearance and the disappearance of that lateral voids during test are then responsible for the jagged experimental resistance vs time curves before failure [18] and compensate the resistance decrease due to the accumulated matter. This effect has never been observed in Structure 1. The voids are located in the half part of the line near the cathode and especially at the junction line/current access line,

which results then in very high local current densities and high local temperatures inducing a very important flux divergence at that points and resulting in a very high value of the parameters n and m. The failure sites are then located very close to the current access line as shown by the Figure 6.

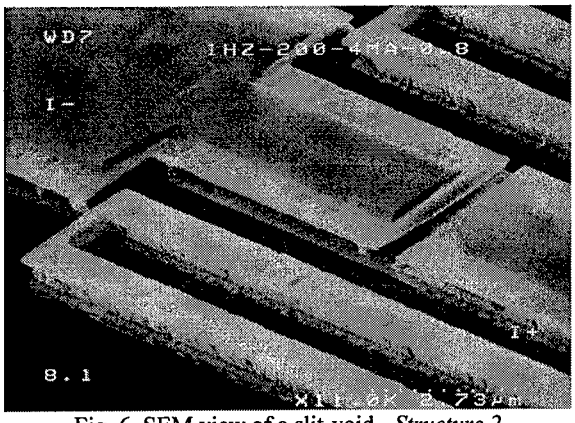



Fig. 6. SEM view of a slit-void - *Structure 2*

For Structure 2, the wafers were submitted, after etching with a standard metal etch process using a $BCl_3$+$Cl_2$ plasma, to a further resist stripping step before D. I. water rinsing. This stage consists in a corrosion passivation with a $H_2O$+$O_2$ plasma at 250°C (Wet D.S.Q) without breaking the vacuum after etching. Afterwards, both samples are submitted to a conventional ashing process using a $H_2O$+$O_2$ plasma at 250°C, and then to a chemical wet cleaning step using EKC 265 product. It has been shown [19] that the residual chlorine after etching was only present on the side wall of the metal lines. Furthermore, the wet D.S.Q and the ashing process induce the formation of a $Al_2O_3$ layer at the lateral metal surfaces which traps the chlorine. This layer is then thicker for Structure 2 than for Structure 1. Corrosion due to the chlorine cannot be responsible for the observed voiding because this effect has not been seen with the same amplitude on the extrusion monitors. It is also reported that the wet cleaning step with EKC 265 reduces the thickness of the $Al_2O_3$ layer.

We suppose that the metal oxide layer is removed, releasing the trapped chlorine, for Structure 1. This is not the case for Structure 2, and the chlorine could be responsible for a worse mechanical link between the metal line and the passivation or could induce a fast diffusion path allowing aluminium grains to be rapidly drifted away at the AlCu/Ti interface (See Fig. 7).

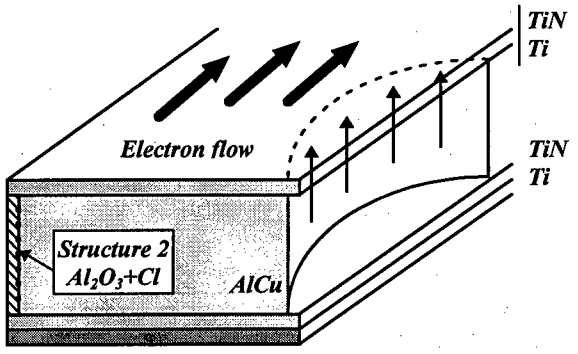



Fig. 7. Lateral voiding for *Structure 2*

## 5. Conclusion

The Black's parameters and their narrow confidence intervals have been extracted from DC and PDC tests using a globalized statistical approach. For these single level AlCu metallizations, the PDC parameters are almost equal to DC ones and the activation energies are the same. Thus, we suppose that the same diffusion mechanism is operating under each stress condition. Furthermore, we found $n_{pdc}$=$m_{pdc}$ at 10kHz, which is in good agreement with the ACM model, and $n_{pdc}$=$2m_{pdc}$ at 1Hz, leading to the On-time model by assuming that the duty cycle accelerating factor is affected by thermal effects. Microstructural characterization has shown large hillocks for each structure and has emphasized the influence of the resist stripping stage after etching on the lateral interface and on the EM behaviour of the test structures, even if the same activation energy were found.

## Acknowledgements

The authors would like to thank M. Heitzmann and D. Louis for the information about the cleaning process.

This work has been carried out within the GRESSI consortium between CEA-LETI and France Telecom CNET and was performed within the framework of the European project ACE.

## References


[1] J.R. Black, IEEE Trans. Electron Devices, ED-16, 1969, p. 338.
[2] L. Brooke, IEEE/IRPS, 1987, p. 136-139.
[3] X. Gui et al, J. Appl. Phys., 80 (9), 1996, p. 4948-4951.
[4] D.G. Pierce et al, IEEE IRPS, 1994, p. 198-206.
[5] P. Waltz et al, Proceedings of the ESSDERC Conf., 1997, p. 484-487.
[6] H. Schafft, J. A. Lechner, IEEE IRPS, 1988, p. 192-202.
[7] J. M. Towner, IEEE IRPS, 1990, p. 100-105.
[8] G. Saporta, « Probabilités, analyse des données et statistiques », Paris, edition Technip.
[9] C. K. Hu et al, IBM J. Res. Develop., 39 (4), 1995, p. 465-497.
[10] D.S. Chhabra, N.G. Ainslie, IBM Technical Report 22 419, July 1967.
[11] A. T. English, E. Kinsbron, J. Appl. Phys., 54 (1), 1983, p. 275-280.
[12] J. A. Maiz, IEEE IRPS, 1989, p. 220-228.
[13] P. Waltz et al, Proceedings of the ESSDERC Conf., 1997, p. 456-459.
[14] P. Waltz et al, Microelectron. Reliab., 37 (10/11), 1997, p. 1553-1556.
[15] W. Leiberg et al, Proceedings of VMIC, 1997, p.467-472.
[16] R. A. Augur et al, AIP Conf. Proc., 373, 1996, p. 279-292.
[17] T. Kouno et al, J. Vac. Sci. Technol., B15 (5), 1997, p. 1800-1804.
[18] H. Onoda et al, J. Appl. Phys., 77 (2), 1995, p. 885-892.
[19] D. Louis et al, Micro and Nanoengineering, PE 1-2, 1997.



PERGAMON

Microelectronics Reliability 38 (1998) 1047–1050

MICROELECTRONICS RELIABILITY

# Mechanical Stress Evolution and the Blech Length: 2D Simulation of Early Electromigration Effects

V. Petrescu[a], A.J. Mouthaan[a], W. Schoenmaker[b], C. Salm[a]

[a] *MESA Research Institute, University of Twente Dept. of Electrical Engineering P.O. Box 217 7500 AE Enschede, The Netherlands*

[b] *IMEC, ASP-TCAD Division, Leuven, Belgium*

## Abstract:

Modeling of stress and electromigration at the microscopic level, in confined interconnect metallic lines with tungsten studs, can very well account for the resistance behaviour in time. The resistance change at saturation for a metallic line with blocking boundaries at both ends can be related, according to the model, to threshold product $(jL)_c$ found by Blech [1].

© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

The problem of electromigration has been around in microelectronics from the early sixties, but so far a complete physical understanding is still lacking. Electromigration is the drift of atoms under the influence of strong electron winds. With the advent of the miniaturised integrated circuits, the current densities in the metal stripes on a chip can attain huge values (MA/cm2). The usual way to obtain reliable circuits is to avoid such large current densities, by having stripe widths of sufficiently large size. Unfortunately, the ever decrease device sizes may ultimately prevent one from maintaining the design rules where a minimum strip width is always present. In other words, the devices may become so small that the wires and vias are simply too crude for being attached to a single device terminal. Defects due to electromigration may show up in actual circuits after long operation times. In order to obtain predictions for product life times within the time-to-market constraints, accelerated tests are performed. However, in order to relate the outcomes of these tests to actual field performance, a good understanding of electromigration is a necessity. We will argue that a simulation tool can contribute considerably to a better understanding of electromigration.

Early effects in electromigration are resistance change and the evolution of mechanical

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00105-X

stress in grains and regions of grain boundaries (called cluster regions). It is assumed that with a realistic value of the concentration of vacancies in the metallisation, the redistribution of these vacancies due to the electron wind force causes large stress gradients. These stress changes induce a piezoresistive effect, changing the resistance of a line. A 2D simulation has been carried out using this model that shows the characteristic Blech length: the critical length of a grain in which a mechanical stress gradient and diffusion force can still be built up that will balance the electron wind force. Lines (or grains) shorter that this length are 'electromigration-hard'.

## 2. Modeling and simulation of electromigration

The modeling is based on the idea that vacancies are physical entities that obey balance laws modified with sources and sinks. In a vacancy electromigration formulation [2], creation/annihilation of vacancies at sinks and sources is the main factor that gives rise to electromigration-induced stresses when confined interconnect lines are in discussion. The models of electromigration are between the atomic modeling and the semi-empirical modeling, i.e. vacancy fluxes and stress migration is described in a set of coupled partial differential equations which expresses the underlying balances of matter and force fields. In the model, based on Clement's and Thompson's model [3], the resistivity of the line is related to the stress evolution according to the piezoresistive effect [4]. For the vacancy fluxes, we have the following balance equation:

$$\frac{\partial C}{\partial t} = -\nabla J_v + G \qquad (1)$$

where $J_v$ is the flux of vacancies, C is relative vacancy concentration ($C=C_v/C_L$, where $C_v$ is the vacancy concentration and $C_L$ is the lattice sites concentration) and G is a source/sink term describing the annihilation and creation of vacancies. The vacancy current is composed of a drift and diffusion current:

$$J_v = -D_v \nabla C + \frac{D_v Z^* e}{kT} CE \qquad (2)$$

The relative vacancy concentration is related to the local hydrostatic stress:

$$C = C_o \exp\left(\frac{\sigma\Omega}{kT}\right) \qquad (3)$$

where $C_o$ is the initial relative vacancy concentration, $\sigma$ the mechanical stress and $\Omega$ the atomic volume. The stress and the concentration of lattice sites ($C_L$) are related according to Hooke's law:

$$G = \frac{\partial C_L}{\partial t} = -\frac{C_L}{B}\frac{\partial \sigma}{\partial t} \qquad (4)$$

The piezoresistive effect relates the change in the resistivity due to change in stress [4]:

$$\frac{1}{\rho}\frac{\partial \rho}{\partial \sigma} = \mathrm{K} \qquad (5)$$

where $\rho$ is the resistivity, $K=1.2\times10^{-5}$ $MPa^{-1}$ is the piezoresistive coefficient for aluminum. Replacing the stress ($\sigma$) from equation (3) in equation (5) we can relate directly the resistivity to the vacancy concentration:

$$\rho = \rho_o \left(\frac{C}{C_o}\right)^{\gamma} \qquad (6)$$

where $\rho_o$ is the initial resistivity.

## 3. Results of the simulation

The two-dimensional segment of an Al metal stripe, of length l, is shown in Fig.1 with a cluster region of length L. The line is between two tungsten vias that act as blocking boundaries (in Fig.1: contact left and contact right). Two different line lengths l have been considered: a short line with l=16 μm (L=8 μm) and a line with l=75 μm line with a 15 μm length cluster region. The width of the line (W) in both cases is 4 μm.

For the short line, two stages can be distinguished in the resistance change evolution (Fig. 2, l=16 μm):

- stage I (Fig. 2), where linear increase in time corresponds to the diffusion of Al atoms/vacancies along grain boundaries and with a lower rate in the bulk (a typical stress distribution is shown in Fig. 3) and a second stage
- stage II (Fig. 2), where resistance saturates.

According to Fig. 2, resistance saturation stage can be identified to the period of no further evolution in mechanical stress at the blocking boundaries. Simulations prove that for a stress difference between stripe ends of $\Delta\sigma = 455 - (-330) = 785$ MPa (Fig. 4), diffusional back force and electromigration force annihilates and a constant stress distribution, in time, at tungsten plugs (Fig. 4) is noticed. At this stage the current density ($j=10^6$ A/cm$^2$) becomes the critical current density and the length l of the line, the critical length ($l_c$=16 μm). From the Blech threshold product [1] $(jl)_c = \Delta\sigma\Omega/(\rho Z^* e)$ ($\rho=5\times10^{-6}$ Ωcm, Z*=10), the back-flow stress $\Delta\sigma$ is 757.39 MPa, value that confirms the simulation results. The resistance saturation and a constant back-stress gradient observed for the short line indicates the suppression of further electromigration damage.

For the longer line, no saturation in resistance was observed (Fig. 2, l=75 μm, stage II) and the increase in resistance (stage I) is larger that in short line. This line can be suspected of serious electromigration damage (void formation, passivation cracking) before the back-stress gradient will balance the electromigration force. A typical stress distribution, corresponding to stage II for the long line is showed in Fig. 5.

Similar experimental results on resistance change evolution have been reported by Filippi [5].

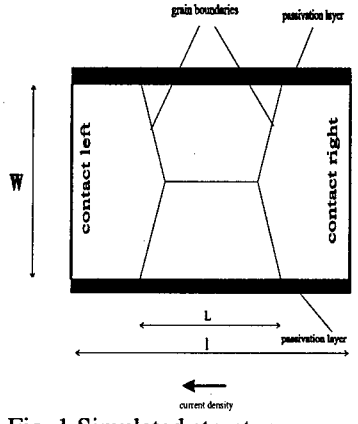



Fig. 1 Simulated structure

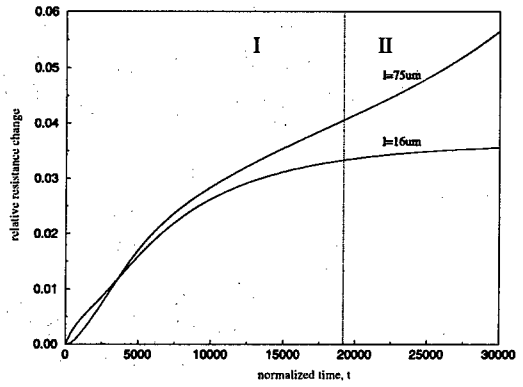



Fig. 2. Relative resistance change

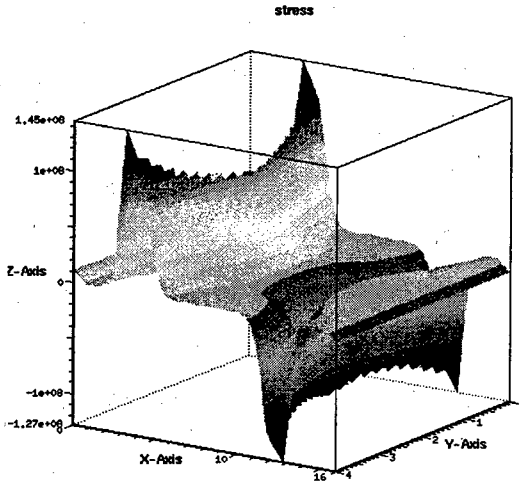



Fig. 3. Stress distribution in [Pa] corresponding to stage I from Fig. 2 for the short line (l=16 μm)

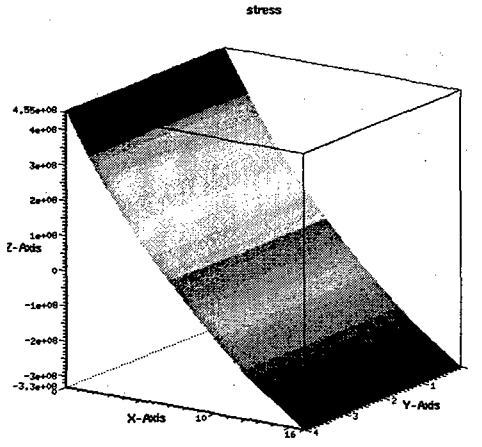



Fig. 4. Stress distribution [in Pa] corresponding to stage II from Fig. 2 for the short line (l = 16 μm)

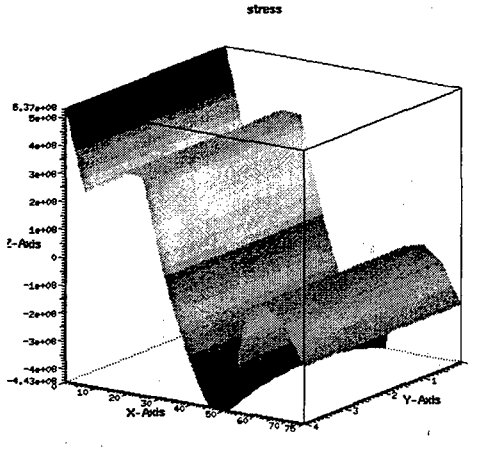



Fig. 5. Stress distribution [in Pa] corresponding to stage II from Fig. 2 for the long line (l = 75 μm)

## 4. Conclusions

The outcomes of the simulations provide an important tool for a better understanding of electromigration and stressmigration phenomena.

## 5. References

[1] I.A. Blech, J. Appl. Phys. 47, 1976, pp 1203.
[2] M.A. Korhonen, S. Rzepka, R.G. Filippi, C.-Y. Li, Proc. of the Fourth Int. Workshop in Stress Induced Phenomena in Letallization, 1997, pp 303.
[3] J.J. Clement, C.V. Thompson, J. Appl. Phys. 78 (2), 1995, pp. 900.
[4] N.L. Beverly, G.B. Alers, J.A. Prybyla, Appl. Phys. Lett. 68 (17), 1996, pp. 2372.
5. R.G. Filippi, G.A. Biery, R.A. Wachnik, J. Appl. Phys. 78 (6), 1995, pp. 3756.

PERGAMON

Microelectronics Reliability 38 (1998) 1051–1056

MICROELECTRONICS RELIABILITY

# Systematic Derivation of Latchup Design Rules for Submicron CMOS Processes from Test Structures

J.A. van der Pol [a], P.B.M. Wolbert [b]

[a] *Waferfab AN,* [b] *Q&R Department, Consumer Systems Nijmegen, Email: Jacob.vanderPol@nym.sc.philips.com Philips Semiconductors, Gerstweg 2, 6534 AE Nijmegen, The Netherlands*

**Abstract**

A dedicated set of test structures can be used to determine the latchup susceptibility of a process as a function of $n^+p^+$-spacing, placement of $N_{well}$ and substrate contacts and guardring width and distance to the injecting junction. A systematic approach is given for the translation of these test structure data into latchup design rules resulting in latchup robust products. The methodology is demonstrated using data from a submicron CMOS process on $p^-/p^{++}$ epitaxial substrates. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Latchup is an intrinsic reliability risk for CMOS processes [1,2] due to the presence of built-in parasitic thyristors in this technology. In case of a p-substrate, the thyristors consist of lateral NPN and vertical PNP transistors, see fig.1.

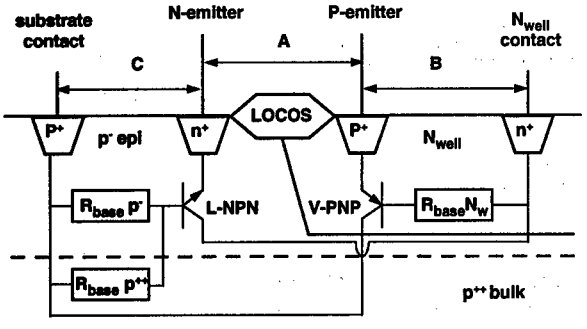



Fig. 1 : Schematic view of a latchup test structure showing the parasitic bipolar transistors and base resistances, the relevant design rules (A: $n^+p^+$-spacing, B: distance N-well contact to $P^+$ emitter and C: distance $P^+$ substrate contact to $N^+$-emitter) and the N- and P-emitters.

As for new generation submicron processes the $n^+p^+$-spacing (that determines the base-width and thus the gain of the parasitic bipolar transistors) continues to shrink, the latchup susceptibility increases strongly. In principle, latchup free products can be obtained by ensuring that the thyristor holding voltage $V_h$ is larger than the supply voltage $V_{dd}$. This appears however to be impractical for submicron processes as $V_h > V_{dd}$ is only achieved for $n^+p^+$ spacings significantly larger than the minimum design rules, even if latchup robust $p^-/p^{++}$ epitaxial substrates are used [3,4]. This is illustrated by the data in fig.2 from two 5V 0.7μm twin well CMOS processes from two waferfabs using 10 Ωcm $p^-$ epi on a 0.01Ωcm $p^{++}$ substrate. Process A has a significantly higher temperature budget than process B1 (the alignment markers are made by a double LOCOS oxidation instead of by a silicon dry etch) which results in an about 1.5μm thinner remaining epi thickness at the end of the process for process A due to diffusion of the $p^{++}$ bulk into the $p^-$ epi layer. This explains that the 9μm epi $V_h$ data of process A correspond to those of 7.5μm epi of process B1. Fig. 2 shows that by going to very thin epi layers, $V_h$ approa-

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00145-0

ches $V_{dd}$ but this option is limited by $P^+/N_{well}/p^{++}$ substrate punch-through and $R_{sheet}$ $N_{well}$ requirements, see e.g. fig. 3. Thus there is a clear need for proper design rules to obtain latchup robust products. In this paper a systematic method is proposed for the derivation of those design rules from test structures. This method is illustrated using data from one of the above CMOS processes on epitaxial substrates. However, it is also applicable for processes on $p^-$ bulk substrates with or without buried layers.

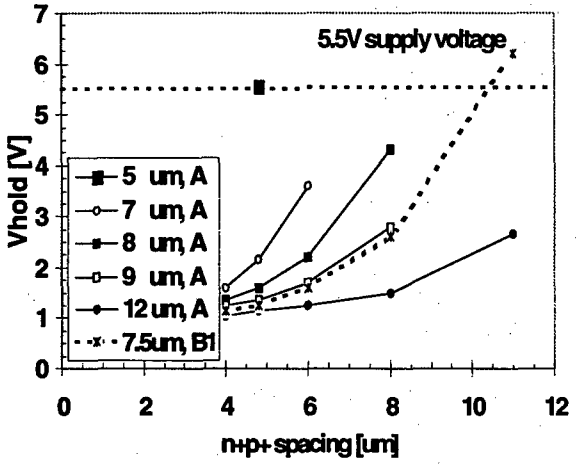



Fig. 2 : Holding voltage at 125°C as a function of $n^+p^+$ spacing A for various epitaxial layer thicknesses and two different 0.7μm processes (A and B1) from two waferfabs.

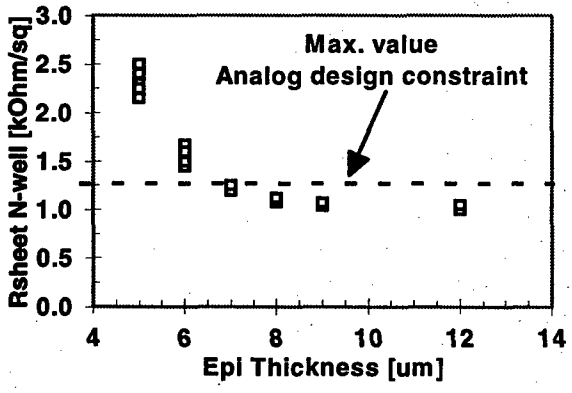



Fig. 3 : $N_{well}$ sheet resistance as a function of epitaxial layer thickness for process A. For small thicknesses the $p^{++}$ doped substrate overdopes the $N_{well}$ resulting in an increase of the sheet resistance. Analog design constraints determine the maximum allowable $R_{sheet}$ $N_{well}$.

## 2. Latchup susceptibility reduction options

Facing the fact that $V_h<V_{dd}$, there are basically only three options to improve latchup robustness. First, one can increase the thyristor P- and N-emitter trigger currents $I_{trig,p/n}$ (hole and electron injection respectively) by reducing the gain of the parasitic bipolar transistors and/or their base resistances by means of process as well as design options, see table 1. Second, one can reduce the number of carriers reaching the thyristor after injection at the product I/O-bondpads by applying $P^+$ or $N^+/N_{well}$ guardrings, see e.g. fig.4.

| Process option | Design option | Effect on gain bipolar transistors | Effect on base resistance |
|---|---|---|---|
| $p^-$ epi layer on $p^{++}$ substrate | | - | Yes (LNPN only) |
| N / $P_{well}$ dose | | Yes | Yes |
| $n^{++}/p^{++}$ buried layers | | Yes (VPNP only) | Yes |
| Silicidation | | Yes | - |
| | $n^+p^+$-spacing | Yes | - |
| | $n^+p^+$-spacing partitioning | Yes | - |
| | Placement of $N_{well}$/substrate contacts | - | Yes |

Table 1: Impact of process and design options on gain and base resistance of LNPN and VPNP parasitic bipolar transistors, assuming a $p^-$-type substrate

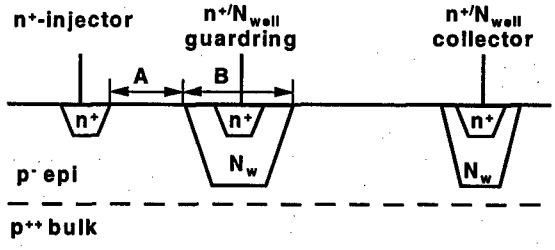



Fig. 4 : Schematic view of a $N^+/N_{well}$ guardring test structure showing the relevant design rules (A: distance guardring to $N^+$-injector and B: guardring width)

Finally the injection points can be placed at a larger distance *d* from the thyristors which also reduces the current density at the location of the thyristors due to geometrical spreading of the injected carriers. In case of an injecting diffusion with length *L*, the decrea-

se of the current density with distance $d$ has been calculated for the two extreme cases shown in fig.5a and 5b. Assuming a 2D-spreading of the injected current, we derive the equations (1) and (2) respectively for the current density reduction factors $F_{spread}$. Note that this a worst case approach as in reality near-3D spreading will occur, resulting in even lower values of $F_{spread}$.

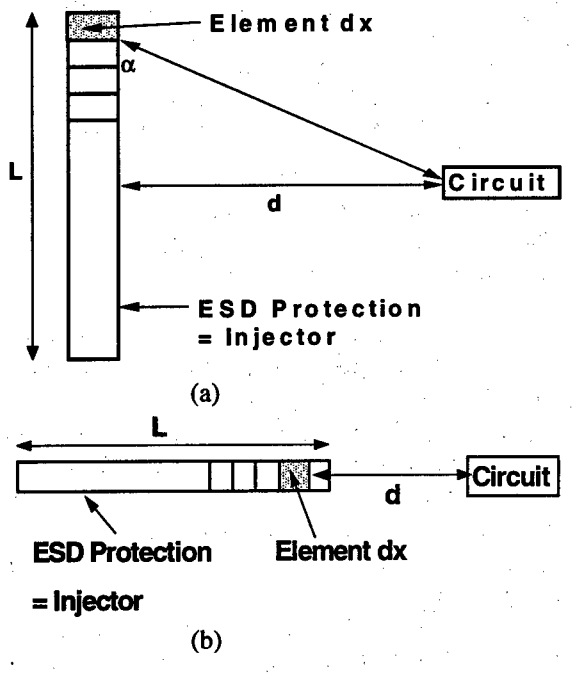



Fig. 5 : Latchup sensitive circuit located a) perpendicular and b) parallel at a distance $d$ from a uniformly injecting diffusion with length $L$ (e.g. an ESD protection).

Circuit located perpendicular to injector with length $L$:

$$F_{spread} = \frac{1}{\pi}\left(\ln\left(\tan\left(\frac{\pi}{4}\right)\right) - \ln\left(\tan\left(\frac{\arctan\left(\frac{2d}{L}\right)}{2}\right)\right)\right) \quad (1)$$

Circuit located parallel to injector with length $L$:

$$F_{spread} = \frac{1}{2\cdot\pi}\left(\ln\left(\frac{d+L}{d}\right)\right) \quad (2)$$

Fig.6 shows some numerical values for an injector with length $L$=100µm (typical length of an ESD protection that acts as injecting diffusion). We see that are only minor differences between the extreme cases for longer distances d and that the current density decrease in general can be reasonable approximated to be proportionally to ≈ln(1+$L/d$).

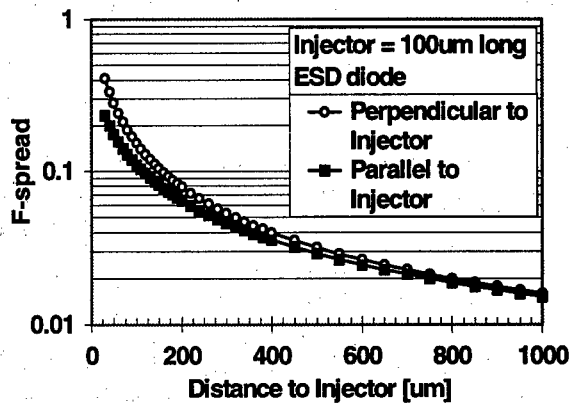



Fig. 6 : Current density reductions factors $F_{spread}$ as a function of distance $d$ and orientation to a uniformly injecting diffusion with length $L$ equal to 100µm.

### 3. Design rule derivation approach

In order to ensure that a product can withstand a specified injection current $I_{injection}$ from the product bondpads, the following criterion must be satisfied :

$$\left|J_{trig,p/n}\right| \geq \left|\frac{I_{injection}}{L_{injection}} \cdot F_{escape} \cdot F_{spread}\right| \quad [\mu A/\mu m] \quad (3)$$

with $J_{trig,p/n}$ the trigger current densitie of the thyristor in case of hole and electron injection respectively, $L_{injection}$ the perimeter of the injecting junction(s) connected to the bondpad (typically the ESD-protection), $F_{escape,p/n}$ the fraction of injected carriers that 'escaped' from the guardrings and $F_{spread}$ the current density reduction factor due to the geometrical spreading of the injected current. Equation (3) holds for both positive and negative pulses (hole injection from P-emitter and electron injection from N-emitter respectively).Typical values for $I_{injection}$ and $L_{injection}$ are +/- 100mA (JEDEC latchup qualification requirement) and 200µm (perimeter typical ESD protection) respectively resulting in a maximum injected current density at the bondpad of ≈500µA/µm. Using guardring efficiency and thyristor trigger currents data obtained from latchup test structures, see fig.7 and 8, one can now determine what the maximum allowed distances of $N_{well}$ and $P^+$ substrate contacts to the emitter diffusions are at minimum $n^+p^+$ spacing and what the guardring width and distance re-

quirements are. Note that here data obtained at maximum junction temperature must be used as latchup trigger currents are lowest at maximum temperature [1,2,5]. Thus the appropriate design rules can be derived as will be illustrated using data from the CMOS process B1, see section 4.3.

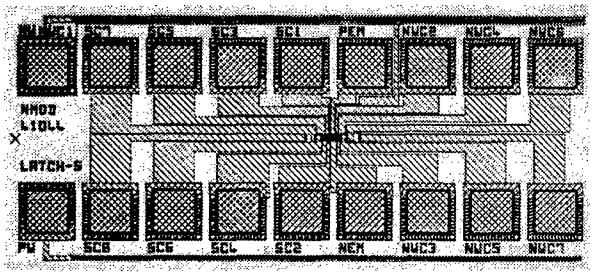


Fig. 7 : Schematic view of a latchup test structure to determine effect of $N_{well}$ and $P^+$ substrate contact placement on latchup sensitivity.

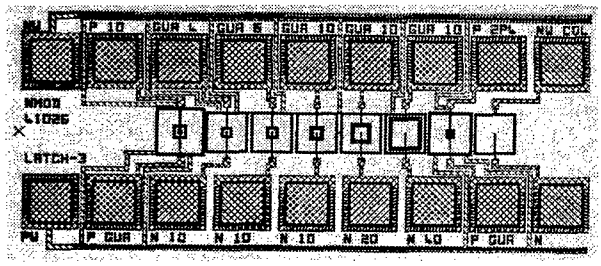


Fig. 8 : Schematic view of a test structure to determine effect of $N^+/N_{well}$ guardring width and distance to N-emitter on the electron collection efficiency.

## 4. Application: design rule derivation for a CMOS process on $p^-/p^{++}$ epitaxial substrates

### *4.1. Impact $P^+$ substrate contact placement and $P^+$ guardrings*

In case of $p^-/p^{++}$ epi substrates, the $p^{++}$ bulk acts as a low ohmic shunt for the base resistance of the lateral parasitic NPN-transistor otherwise formed by the high ohmic $p^-$ epi layer, see fig.1. This strongly improves P-emitter trigger currents as shown in fig.9, particularly for $n^+p^+$ spacings larger than the epi layer thickness. In that case the injected hole current flow changes from lateral through the $p^-$ epi layer to vertical through the $p^{++}$ bulk.

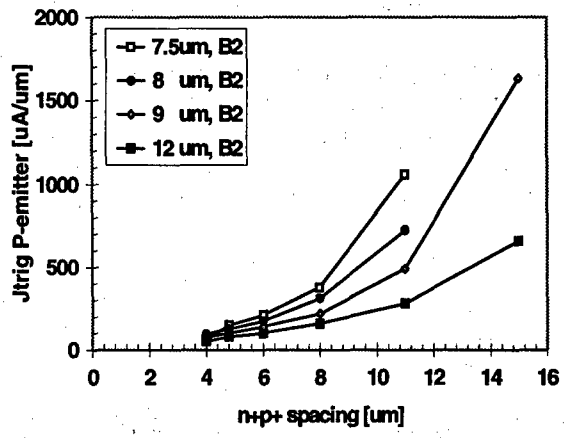



Fig. 9 : P-emitter trigger current of process B2 at 125°C as a function of $n^+p^+$-spacing A (see fig.1) for various epitaxial layer thicknesses.

Fig.10 shows that the $P^+$ substrate contact placement is not critical. The $p^{++}$ bulk namely acts as a kind of equipotential surface that sinks all the injected holes and redistributes them over all available substrate contacts in the layout, see fig.11. The resistance $R_1$ from a $P^+$ contact through the $P_{well}$ and epi layer is in the order of a few kΩ while the substrate spreading resistances $R_{sub1,2}$ are less than 100Ω for distances $d_2$ less than 1 mm. As a result, $F_{spread}$ is very small (<<0.1). For the same reason the $P^+$ guardring efficiency is fully determined by the ratio of the $P^+$ guardring area versus the total $P^+$ area in the layout and thus the $P^+$ guardring width and distance to the $P^+$ emitter are not very critical. The thinner the epi layer, the more effective the $p^{++}$ base shunt and its function as a hole current sink.

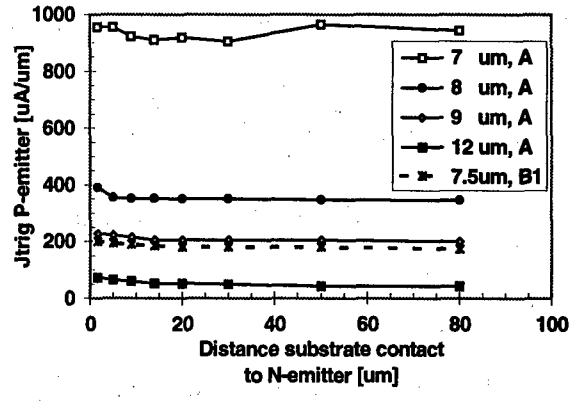



Fig. 10 : P-emitter trigger current at 125°C of process A and B1 as a function of $P^+$ substrate contact to N-emitter distance C (see fig.1) at 4.8μm $n^+p^+$-spacing for various epi layer thicknesses.

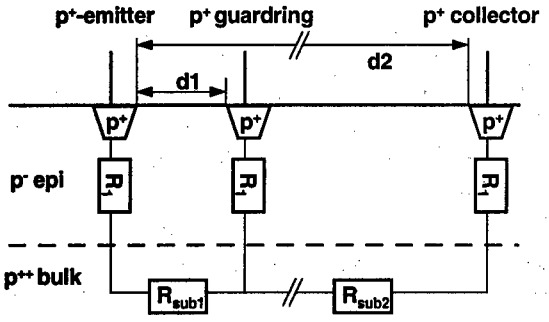



Fig. 11 : Schematic view of the distribution of injected holes over the various $P^+$ substrate contacts.

### 4.2. *Impact $N_{well}$ contact placement and $N^+/N_{well}$ guardrings*

Another feature of $p^-/p^{++}$ epi substrates is that the injected electrons are confined to the epi layer due to the build-in potential between the $p^-$ epi and $p^{++}$ substrate and the small minority carrier diffusion length (≈1μm) in the $p^{++}$ substrate. First, this results in a slight increase of the ($N_{well}$) base resistance of the parasitic PNP transistor and thus actually in a somewhat reduced N-emitter trigger current for thinner epi layers and small $n^+p^+$ spacings, see fig.12.

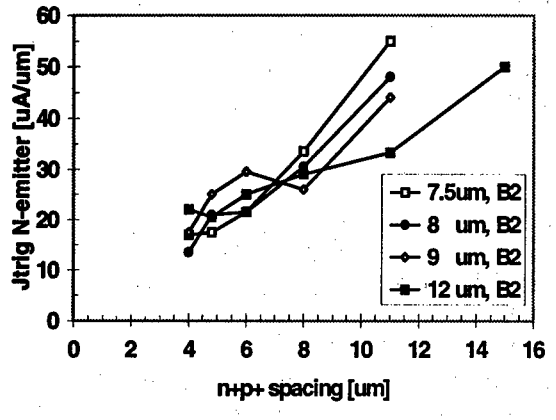



Fig.12: N-emitter trigger current at 125°C of process B2 as a function of $n^+p^+$-spacing A (see fig.1) for various epitaxial layer thicknesses.

Second, $N^+/N_{well}$ guardrings become very efficient as due to the injected electron confinement in the epi layer the distance of the $N^+/N_{well}$ guardring to the injector is irrelevant, see fig.13. Furthermore fig.14 shows that the collection efficiency improves logarithmically with the guardring width. This can be understood using fig.15, where we divided the $p^-$ epi layer between $N_{well}$ bottom and $p^{++}$ bulk in squares. The probability that an electron on its 'random walk' diffusion path does pass such a square is about $1/e$ ($e$=2.7) because the electron recombines or is collected as soon as it hits the $p^{++}$ bulk or $N_{well}$ respectively. The number of escaped electrons does thus decrease exponentially with the number of squares and thus guardring width. Also here thinner epi layers are beneficial as this improves the electron confinement (i.e. increases the number of squares for a constant guardring width).

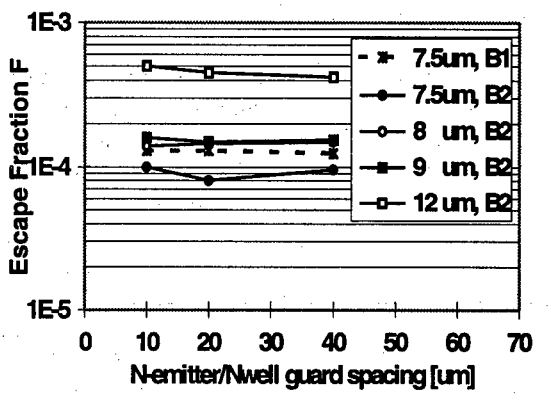



Fig.13 : N+/Nwell guardring efficiency at 125°C of process B1 and B2 as a function of distance A (see fig.4) between $N^+$ injector and a 10μm wide $N^+/N_{well}$ guardring for various epi layer thicknesses.

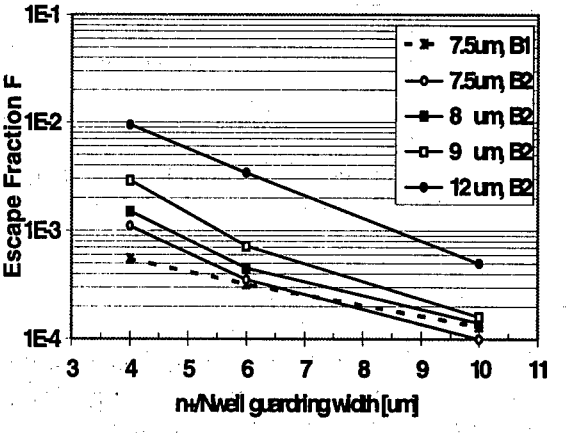



Fig.14 : N+/Nwell guardring efficiency at 125°C of process B1 and B2 as a function of guardring width B (see fig.4) for various epi layer thicknesses.

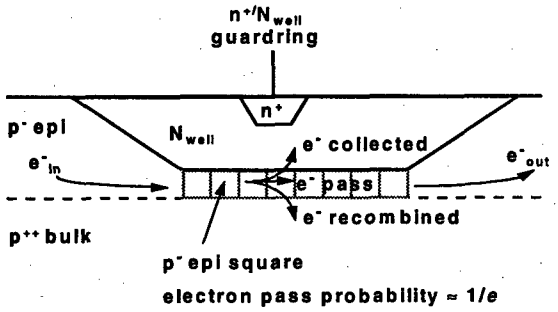



Fig.15 : Escape fraction of electrons decreases exponentially with the number of $p^-$ epi squares and thus $N^+/N_{well}$ guardring width.

Finally fig.16 shows the effect of the placement of $N_{well}$ contacts on the N-emitter trigger current for various epi layer thicknesses. As the $N_{well}$ contact to P-emitter spacing directly determines the base resistance of parasitic PNP transistor, it has a strong effect on the trigger current.

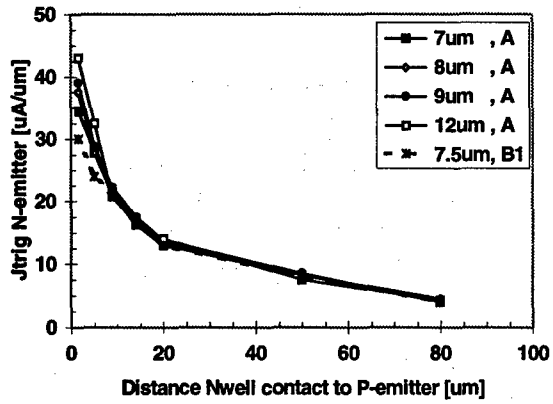



Fig.16: N-emitter trigger current at 125°C of process A and B1 as a function of $N_{well}$ contact to P-emitter distance B (see fig.1) at 4.8μm $n^+p^+$-spacing for various epi layer thicknesses.

### *4.3 Design rules for process B1*

Design rules can now be easily derived from the above data where we take the data of process B1 as example. In case of hole injection (positive trigger currents), fig.10 shows that for process B1 with a 7.5μm $p^-$ epi layer $J_{trig,p}$= 190μA/μm at the 4.8μm minimum $n^+p^+$ spacing. As $F_{spread}$<<0.1, see section 4.1, the criterion in equation (3) is thus easily met and the substrate contact and $P^+$ guardring design rules can be very relaxed (e.g. one substrate contact for every 200μm and use of minimum $p^+$-width for the $P^+$ guardring). In case of electron injection (negative trigger currents), fig.13 shows that for process B1 with a 7.5μm $p^-$ epi layer, a 4μm wide $N^+/N_{well}$ guardring results in $F_{escape}$ =5.5·10$^{-4}$. Assuming $F_{spread}$=1 (very conservative, see fig.6) we find that $J_{trig,n}$ should be ≥500x5.5·10$^{-4}$x1= 0.28μA/μm. One now can determine the maximum allowable $N_{well}$ contact to P-emitter spacing from fig.14. Using a design rule of 100μm, we find $J_{trig,n}$= 3.5μA/μm, which still provides a factor 12 safety margin. The above demonstrates that CMOS processes on epi substrates are very latchup robust provided proper design rules are used.

## 5. Conclusions

Using a dedicated set of test structures, the latchup susceptibility of a number of submicron CMOS processes on $p^-/p^{++}$ epitaxial substrates have been characterized as a function of $n^+p^+$-spacing, placement of $N_{well}$ and substrate contacts and guardring width and distance to the injecting junction. Subsequently it has been shown how these data can be translated into latchup design rules taking into account the geometrical spreading of the injected carriers. It is demonstrated that this approach results in very latchup robust products in case of $p^-/p^{++}$ epitaxial substrates. The method is also applicable to processes on non-epitaxial substrates.

## References

[1] R.R. Troutman, 'Latchup in CMOS technology', Kluwer Academic Publishers, Boston, (1986)

[2] E.A. Amerasekera, 'Failure mechanisms in semiconductor devices', Chapter 3, Wiley, Chichester, (1997)

[3] E.A. Amerasekera, 'Designing latchup robustness in a 0.35um technology', Proc. IRPS., pp.280-285, (1994)

[4] M.J. Chen, S.S. Ho, P.N. Tseng, R.Y. Shiue, H.S. Lee, J.H. Chen, J.K. Jeng, Y.N. Jou, 'A compact model of holding voltage for latchup in epitaxial CMOS', Proc. IRPS, pp. 339-345, (1997)

[5] T. Aoki, 'A discussion on the temperature dependence of latchup trigger current in CMOS/BiCMOS structures', IEEE Trans. Elec. Dev., ED-40, pp.2023-2028, (1993)

PERGAMON

Microelectronics Reliability 38 (1998) 1057–1061

MICROELECTRONICS RELIABILITY

# REVERSIBILITY OF CHARGE TRAPPING AND SILC CREATION IN THIN OXIDES AFTER STRESS/ANNEAL CYCLING

P. Riess, R. Kies, G. Ghibaudo, G. Pananakakis and J. Brini

*Laboratoire de Physique des Composants à Semiconducteurs, ENSERG, BP 257, 38016 Grenoble, France*

**Abstract**

The reversibility of charge buildup and SILC generation in thin oxides subjected to successive stress/anneal cycles is investigated. It is demonstrated that in thin oxides both electron trapping and SILC are nearly fully reversible degradation processes having a generation kinetics almost unchanged after several stressing/annealing cycles. The annealing kinetics of the SILC is likely associated to the out diffusion of charged defects (possibly trapped holes or H+) whose characteristics (diffusivity, activation energy) are independent of the oxide thickness. Moreover correlation between electron trapping and SILC generation has been studied.

© 1998 Elsevier Science Ltd. All rights reserved.

## Introduction

The reliability of thin dielectrics is a primary concern for the qualification of newly developed CMOS and EEPROM technologies. Wear out has been recognized for a long time as the main degradation mechanism in thin oxides [1-3]. However, in ultra-thin dielectrics, it is accompanied by a stress induced leakage current (SILC), which prevails in thinner oxides [4-10]. In this work reversibility of charge buildup and SILC generation in thin oxides subjected to successive stress/anneal cycles is investigated.

The devices used in our experiments are large area MOS capacitors grown on $N^+$ active region with 5.5nm, 6.5nm and 7.8nm thick tunnel oxides fabricated by SGS-THOMSON Microelectronics (Agrate, Italy). The test structures have successively

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00107-3

been subjected to a positive Fowler Nordheim electron injection under constant current stress (10mA/cm$^2$) performed at room temperature, and, followed by isothermal annealing at high temperature (300-400°C with duration 0.5-1 hour).

### Stress/annealing of $Q_{ox}$

The trapping properties (oxide charge density $Q_{ox}$ and normalized centroid $\lambda$) have been measured using the DiMaria technique [1] after recording the gate voltages needed for maintaining a constant gate current for both polarities as a function of the injection dose $Q_{inj}$ (varied from $10^{17}$ to $10^{20}$ q/cm$^2$). The SILC creation has been monitored by Ig-Vg measurements after interrupting the stress at particular injection doses.

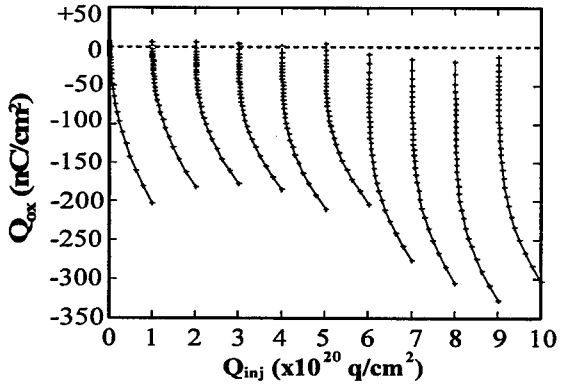



**Fig.1**: Variation of Qox with cumulative injection dose after 10 stress/anneal cycles (tox=7.8nm).

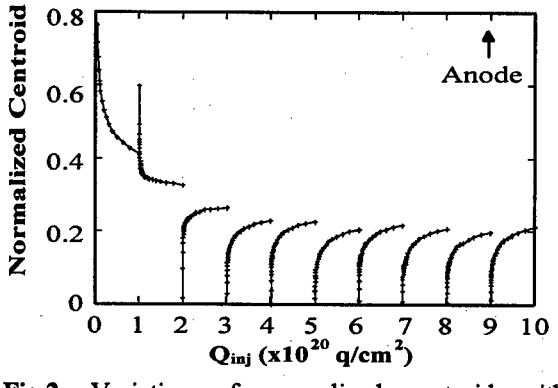



**Fig.2**: Variation of normalized centroid with cumulative injection dose after 10 stress/anneal cycles (tox=7.8nm).

The wear out properties of the oxide as a function of the cumulative injection dose after 10 stress/anneal cycles are reported in Fig. 1. The overall good reversibility of the negative charge buildup can be noticed indicating that the electron trapping after each stress can be almost fully annihilated after high temperature bake. The behavior of the centroid kinetics remains almost unchanged, except for the two first cycles. The difference for the first cycles should be related to the creation of a small positive charge at the early stage of stress which is not reversible. The degradation rate $dQ_{ox}/dQ_{inj}$ is found to follow a power law with the injection dose as $K \times Q_{inj}^{\nu}$ [3], which is weakly modified after several stress/anneal cycles (see Fig. 3). The power law exponent $\nu$ decreases a little, while the trapping efficiency K increases slightly with the cumulative injection dose (see Fig. 4). Similar results have been obtained for 6.5nm and 5.5nm thick oxides with a strong reduction of charge buildup as the oxide thickness is reduced.

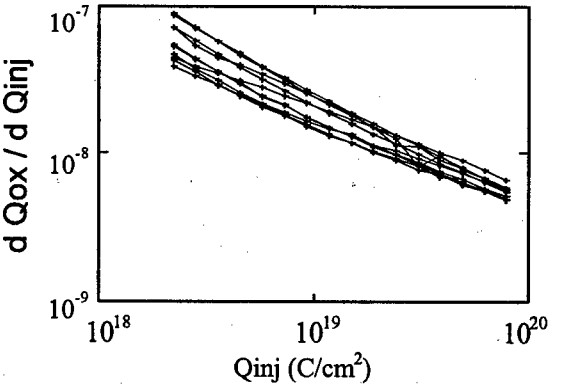



**Fig. 3:** Variation of degradation rate with injection dose after 10 stress/anneal cycles (tox=7.8nm).

To gain more knowledge about the annealing kinetics of $Q_{ox}$ the annealing was done at different temperatures. In this experiment electrons were injected at room temperature, up to a dose of $10^{20}$ e/cm$^2$ with a stress current Jstress = 100mA/cm$^2$. Then the sample was heated up to a temperature T and the evolution of $Q_{ox}$ was measured at constant temperature over $10^4$ seconds. The shift of the Fowler Nordheim voltage was measured at +/- 10 $\mu$A/cm$^2$. The reference of $Q_{ox}$ is not the virgin curve but the curve after maximum injection. As can be seen in Fig. 5 for temperatures below 150°C no annealing is observed. For higher temperatures $Q_{ox}$ becomes positive which means that electrons are detrapped. The detrapping kinetics is more or less

constant for temperatures ranging between 200 °C and 300°C

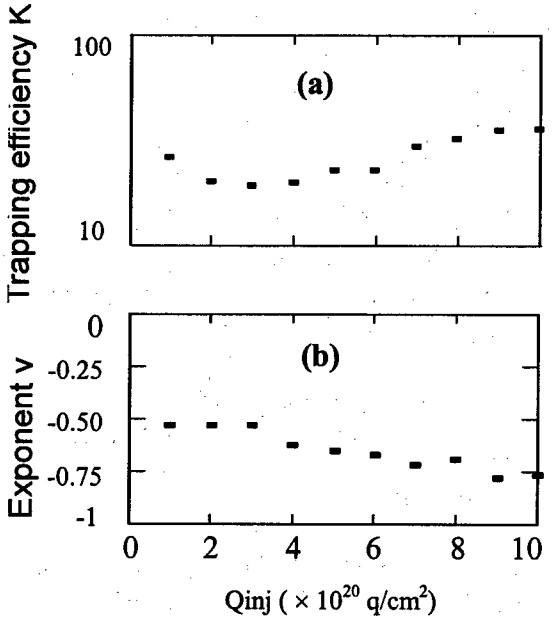



**Fig. 4**: Variation of trapping efficiency K (a) and exponent ν (b) with cumulative injection dose (tox=7.8nm).

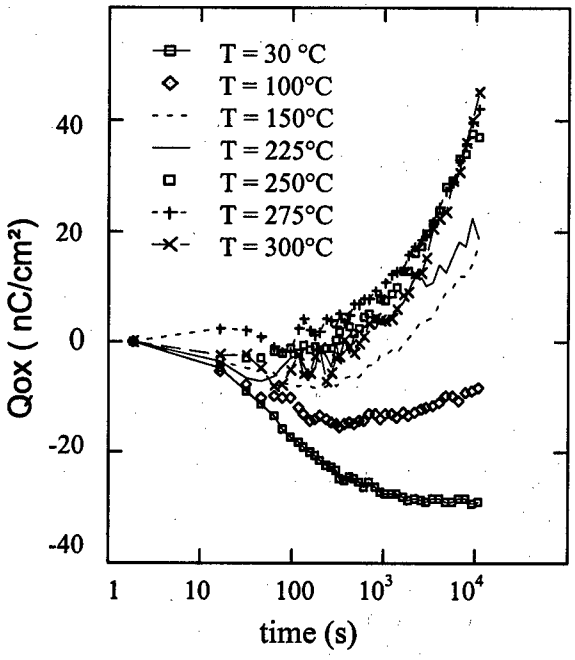



**Fig.5:** Qox annealing kinetics for temperatures varying from 30°C up to 300°C over $10^4$ seconds (tox = 7.8nm).

## Stress/annealing of the SILC

The reversibility of SILC generation has been studied on the thinner oxides (5.5 and 6.5nm) and is reported in Fig. 6 and 7. These figures clearly demonstrate that the SILC is a fully reversible phenomenon after annealing. During each stressing period some leakage paths are created and then almost completely annihilated after a high temperature bake (300°C-0.5hour), letting the oxide recovering its initial state. The logarithmic degradation rate kinetics is found to obey a power law [10], $\partial \mathrm{Ln}(J_S) / \partial Q_{inj} = K Q_{inj}^{-\nu}$, which is not altered by successive stressing/annealing cycles (see Fig. 7). A similar behavior has been obtained for the 6.5nm thick oxide after an annealing at higher temperature (400°C).

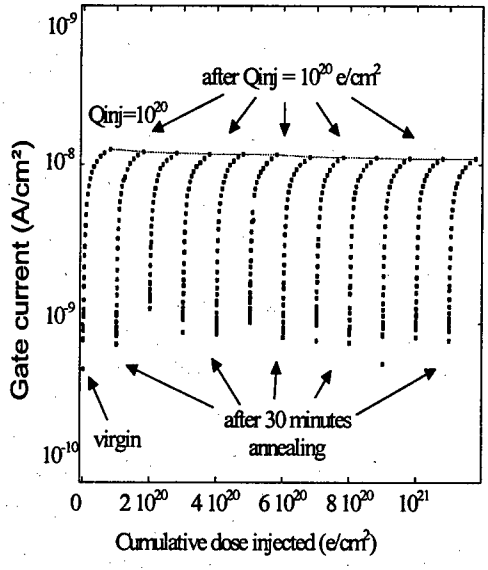



**Fig.6**: Variation of the gate current in the SILC region with the cumulative dose after 12 stressing/annealing cycles (Vg=-4V, tox=5.5nm).

In order to study the SILC annealing kinetics the samples were subjected to a constant current stress up to a dose of $10^{18}$ e/cm$^2$ for stressing temperatures varying from 50°C up to 300°C. Then the samples were annealed at the stressing temperature under 0 bias for $10^4$ seconds and the

SILC recorded against time. Figure 8 shows that for temperatures below 200°C there is no significant

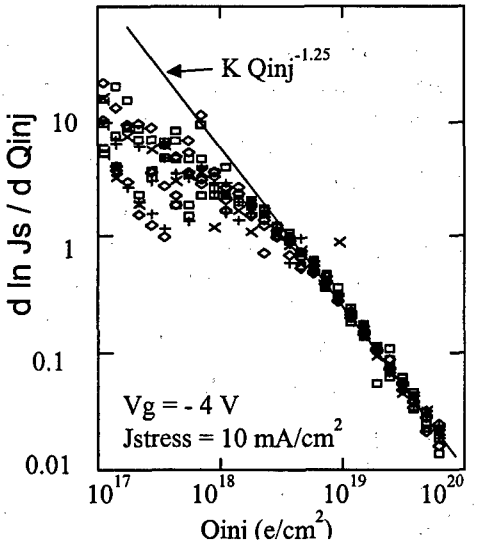



**Fig.7**: Variation of the logarithmic rate of the SILC ∂ ln Js / ∂ Qinj with injection dose after 12 stressing/annealing cycles.

annealing but just a relaxation phase for the SILC with a maximum change of 20%. For higher temperatures an effective annihilation mechanism is progressively activated. Similar experiments under bias show actually that the annealing kinetics is accelerated for both positive and negative polarities. This feature demonstrates that the SILC annealing process is field assisted and by turn that the species at the origin of the SILC are electrically charged. The annealing of the SILC can therefore be interpreted as a diffusion of specific defects out of the oxide. The associated diffusion coefficients deduced from the annealing kinetics time constant have a pre-exponential coefficient D0 ≈ 1 $cm^2/s$ and an activation energy Ea ≈ 1.5±0.1 eV , independently of the oxide thickness.

**Correlation between Qox and SILC ?**

At first glance it can be thought that there is a direct correlation between the SILC and the electron build up. This is supported by the fact that the SILC and the negative charge build up are both reversible and after an annealing at 300°C both the SILC and the negative charge $Q_{ox}$ can be removed (see Fig.9). But if the annealing kinetics of the SILC and $Q_{ox}$ are examined in more detail some discrepancies appear which invalidates the above assumptions.

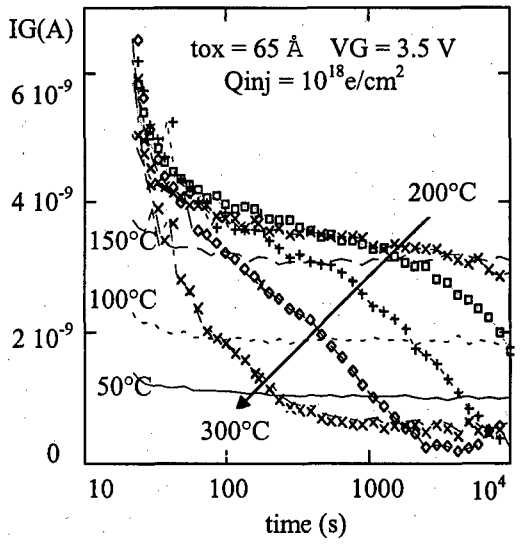



**Fig.8:** SILC annealing kinetics for temperatures varying from 50°C up to 300°C over $10^4$ seconds (tox = 6.5nm).

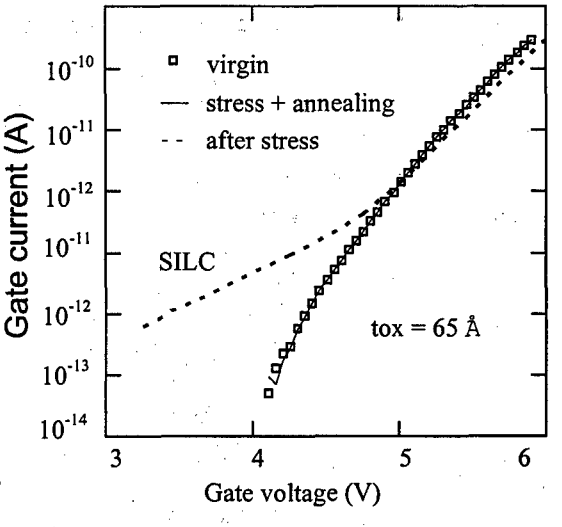



**Fig.9** : Ig-Vg characteristics for a virgin device after stress and after annealing. (tox = 65 Å ).

The question is to know if there is realy a correlation between the recovery of the SILC and the detrapping of electrons. In order to study simultaneously the SILC annealing and the recovery of $Q_{ox}$ an oxide thickness for which both the SILC

and the electron charge buildup are significant was necessary. The 65 Å thick oxide satisfied both of this conditions. The experiment consisted in injecting $10^{20}$ e/cm$^2$ at room temperature and then anneal the sample at higher temperature. During the annealing the evolution of the SILC and the Fowler Nordheim current at high fields were measured for a constant voltage. The surprising aspect is that even when the amplitude of the SILC is decreasing the current at high voltages is constant or even increasing after a while. Once the SILC is totally removed , after a few hundred seconds the current at high fields keeps increasing. This behaviour seems to indicate that there is no direct correlation between the SILC sensed at low voltage and the shift of the Fowler Nordheim current at high voltages attributed to electron trapping. Moreover when the annealing kinetics of the SILC are compared to that of Qox the two kinetics are different. There is negative charge buildup (due to sensing) or only very little annealing of Qox for temperatures below 150°C. For

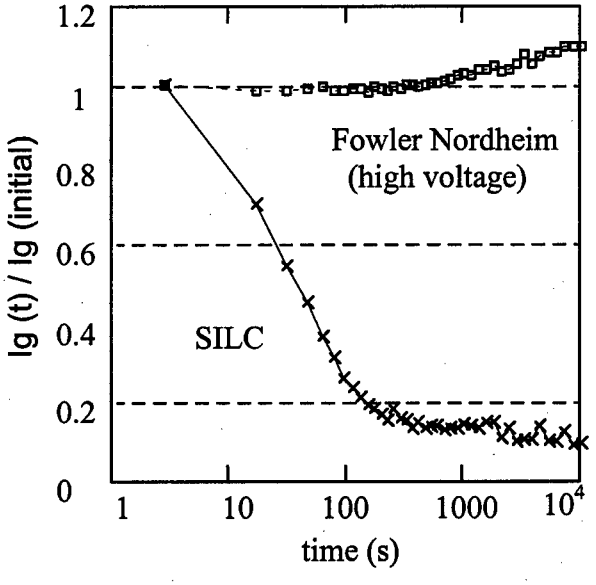



**Fig.10**: Annealing kinetics at high voltage (Fowler Nordheim regime) and low voltage SILC on the same device. (tox = 65 Å, Tanneal = 300°C)

higher temperatures there seems to be no activation of the electron detrapping (see Fig.5). On the other hand the SILC annealing kinetics is enhanced when the temperature is increased (Fig.8). Therefore the origin of the electron trapping and that of the SILC seems to be different. This idea is confirmed if the activation energy of the electron trapping process is compared to that of the SILC generation. The activation energy of the generation of the SILC is relatively small (< 0.05 eV) compared to that of the trapping process ≈ 0.1 eV confirming that the electron trapping and the SILC are not strictly correlated[11-12].

## Conclusion

We have demonstrated that in thin oxides both electron trapping and SILC are nearly reversible degradation processes, and, that their generation kinetics remain almost unchanged after successive stressing/annealing cycles. In addition the annealing kinetics of the SILC can be ascribed to the out diffusion of charged defects (possibly trapped holes or H+) whose parameters (diffusivity, activation energy) are independent of the oxide thickness. Finally there is no evidence for a direct correlation between negative charge buildup (electron trapping) and the SILC as evidenced by their different annealing kinetics and generation activation energy.

1. D. DiMaria, J. Appl. Phys., 47, 4073 (1976).
2. D. DiMaria and D. Arnold, E. Cartier, Appl. Phys. Lett., 61, 2329 (1992).
3. C. Papadas, G. Ghibaudo, F. Pio, C. Montserié, G. Pananakakis, P. Mortini, C. Riva, Sol. State Electron., 37, 495 (1994).
4. P. Olivo, T. Nguyen, B. Ricco, IEEE Trans. on Electron Devices, ED-35, 2259 (1988).
5. R. Rofan an C.Hu, IEEE Electron Dev Lett, EDL-12, 632 (1991).
6. H. Satake and A. Toriumi, Appl. Phys. Lett.,67, 3489 (1995).
7. K. Okada,S.Kawasaki,Y.Hirofuji Proc. of SSDM, Yokohama, Japan (1994).
8. D. DiMaria and E.Cartier, J. Appl. Phys., 78, (1995).
9. J. DeBlauwe et al, IEDM Tech Digest, p. 343 (1996).
10. A. Scarpa et al, Electronics Letters, 33, 1342 (1997).
11. R. Moazzami and C. Hu , IEDM 1992,p 139
12. R. Kies, PhD thesis, INPG, (1997)

PERGAMON

Microelectronics Reliability 38 (1998) 1063–1068

MICROELECTRONICS RELIABILITY

# Precise quantitative evaluation of the hot-carrier induced drain series resistance degradation in LATID-n-MOSFETs

Georg H. Walter [a,b], Werner Weber [a], Ralf Brederlow [a,b], Reinhard Jurk [c], Carsten G. Linnenbank [a], Christian Schlünder [a], Doris Schmitt-Landsiedel [b], and Roland Thewes [a]

[a] *Siemens AG, ZT ME 2, Otto-Hahn-Ring 6, D 81730 Munich, Germany,*
[b] *Technical University of Munich, D 80290 Munich, Germany*
[c] *Siemens AG, HL PI M, Otto-Hahn-Ring 6, D 81730 Munich, Germany*

**Abstract**

A method is presented which allows to distinguish the drain series resistance increase from other mechanisms contributing to the drain current degradation of hot-carrier stressed n-MOSFETs. Devices with different channel lengths but equal damages are used. The different degradation mechanisms are characterized quantitatively and a model for the drain current degradation is presented. For short stress times, the drain current degradation is dominated by series resistance degradation. For long stress times, however, the contribution of the mechanisms attributed to an "equivalent channel length increase" prevails.
© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

In hot-carrier stressed n-MOSFETs with graded junction profiles (LDD, LATID, etc.) a relatively big portion of the hot-carrier induced damage is located above the drain n⁻-region (cf. Fig. 6c)). Thus a hot-carrier induced drain series resistance increase is obtained. It is widely accepted in the literature, that this effect contributes prominently to the drain current degradation in addition to the impact of the damage located above the channel region close to drain (e.g. [1-3]).

In this paper we present a new experimental method which allows to determine hot-carrier induced drain series resistance increase quantitatively and to distinguish between series resistance and channel region related effects. Experimental results obtained by using this method will be discussed and interpreted considering the physical background.

## 2. Device data

The devices investigated originate from a 0.65 µm n-well CMOS process, specifically optimized for analog applications, $t_{ox}$ = 15 nm, n-MOS-LATID and p-MOS-LDD profiles. Spacer length ≈ 250 nm, leading to low substrate current device design with good hot-carrier immunity and high values of $r_{DS}$.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00112-7

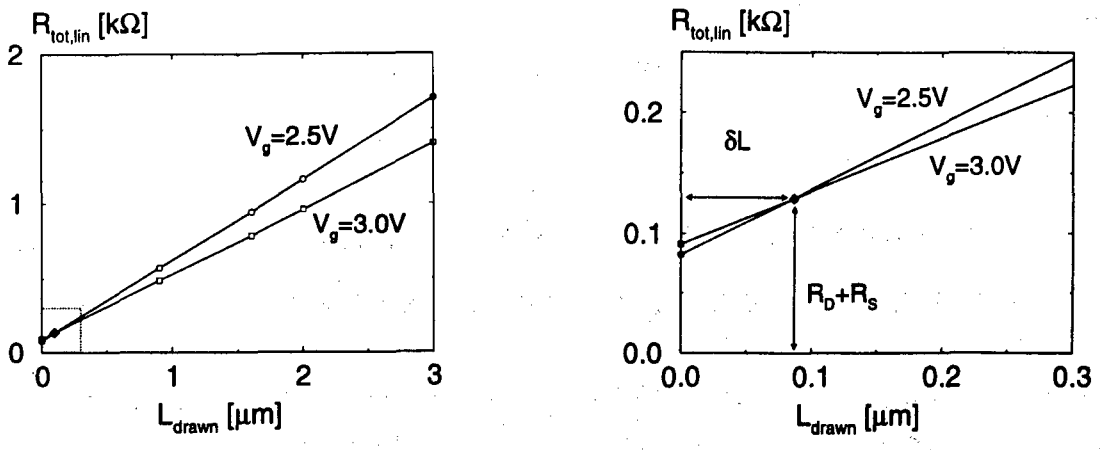



Fig. 1: Characterization method to determine the series resistances of drain and source, $(R_D+R_S)$, and the parameter $\delta L$ (difference between the drawn gate length $L_{drawn}$ and the effective channel length $L_{eff}$) [4, 5]. Left: Plot of the total linear mode transistor resistance $R_{tot,lin} = V_D/I_{D,lin}$ as a function of $L_{drawn}$ for the gate voltages given in the figure. Note: $\Delta V_G = 0.5$ V is chosen here only to demonstrate the method more clearly. Usually, smaller values are used (in Fig. 5: $\Delta V_G = 0.1$ V). Right: Blow-up of left plot featuring the parameters $(R_D+R_S)$ and $\delta L$.
Devices with $W = 10$ µm, further device and process data see text. Characterization drain voltage $V_D = 0.1$ V.

## 3. Method

In unstressed transistors, the sum of the series resistances of drain and source $R_D+R_S$ can be obtained by the following procedure [4, 5]: The total linear mode transistor resistance $R_{tot,lin} = V_D/I_{D,lin}$ of devices with identical properties but different values of the drawn gate length $L_{drawn}$, is plotted as a function of $L_{drawn}$ for slightly different gate voltages $V_{G1}$ and $V_{G2}$ as depicted in Fig. 1. The crossover of the two straight lines achieved for $V_{G1}$ and $V_{G2}$ yields the parameter $(R_D+R_S)$ at the operating point $(V_{G1}+V_{G2})/2$. Furthermore, a correction term $\delta L$ is obtained in this plot, determining the effective channel length $L_{eff} = L_{drawn} - \delta L$.

However, this method is not directly applicable to MOSFETs stressed under usual hot-carrier stress conditions since distribution and amount of damage (for given values of $V_{D,stress}$ and $V_{G,stress}$) depend strongly on the channel length. This problem is solved here using the following procedure:

1.) Devices with $W$ = const. and different values of $L_{drawn}$ are stressed with the multiplication factor under stress $M_{stress} = I_{B,stress}/I_{D,stress}$ held constant at a predefined value for all channel lengths. This is achieved by choosing a constant stress gate voltage and continuously regulating $V_{D,stress}$ during stress. For a given value of $M_{stress}$, the total variations of the stress drain voltage due to regulation and due to varying channel lengths are below ± 2.2 % of $V_{D,stress}$. $V_{G,stress}$ is chosen so that the operating point under stress is always close to $I_{B,stress,max}$ conditions. Since $M_{stress}$ is a measure for the rate of hot carriers and thus also for the lateral electrical field under stress, its channel length-independent constant value guarantees that also the distribution of damage does not depend on $L_{drawn}$.

2.) Characterization of the devices is performed at equal values of the integral

$$Q_{B,stress} = \int I_{B,stress}(t_{stress})\, dt_{stress} \qquad (1)$$

for all channel lengths, and not after given values of stress time as in usual stress experiments. This integral, which is continuously evaluated during stress, is a measure for the amount of damage. Thus, equal values of $Q_{B,stress}$ guarantee channel length independent amounts of damage.

We perform constant base level [6-8] and constant amplitude [6, 9] charge pumping measurements (CPB and CPA) (Figs. 2 and 3) to prove the validity of our approach. The latter is a sensitive monitor for stress-induced interface states in the channel region while the first reveals information on the damage in the n⁻-region (for values of the upper level $V_H$ of the charge pumping pulse below $V_{th}$ in the channel region). For increasing values of $V_H$, also the extension of the monitored area of the n⁻-region increases.

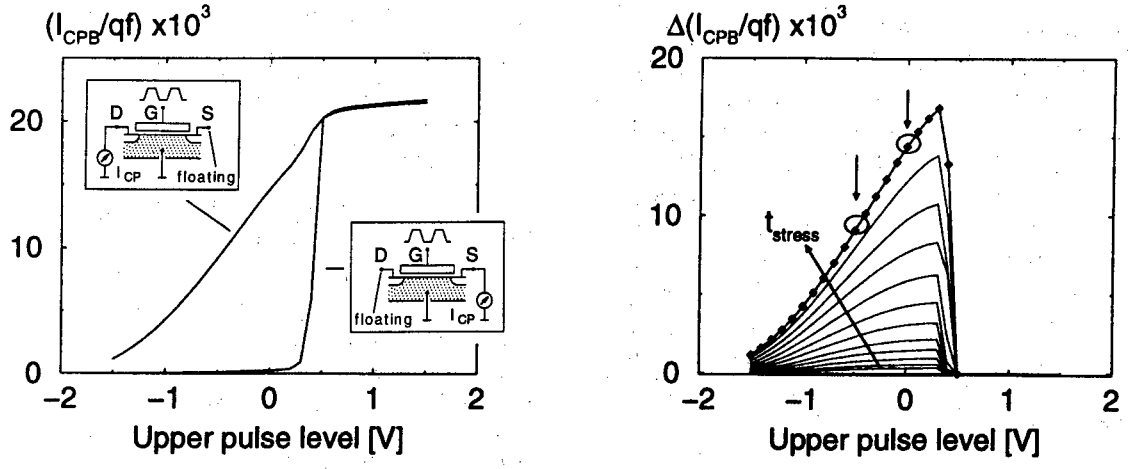



Fig. 2: Example for a constant base level charge pumping result. $W$ = 20 µm, $L_{drawn}$ = 1.5 µm, further process data see text. Characterization parameters: Lower pulse level $V_L$ = - 6 V, upper pulse level $V_H$ = -2 ... +2 V, $t_H$ = $t_L$ = 200 ns, slew rate = $t_{rise}/(V_H-V_L)$ = $t_{fall}/(V_H-V_L)$ = 20 V/µs (c. f. [6-8]). Left: Drain and source side signal after hot-carrier stress. The insets show the configuration of the set-up following [8]. Right: Difference of drain and source side signals after increasing hot-carrier stress. The arrows mark the characterization points $V_H$ = -0.5 V and $V_H$ = 0 V (c.f. Figs. 4 and 8, respectively).

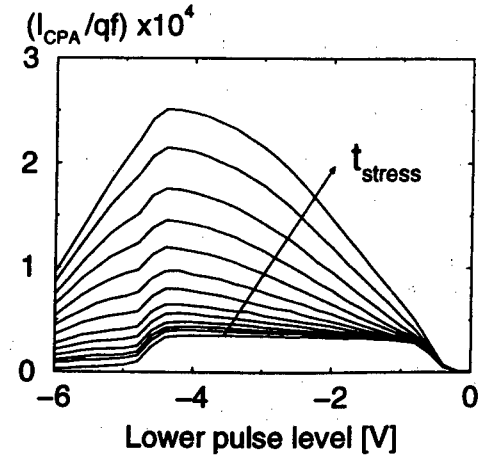



Fig. 3: Example for a constant amplitude charge pumping result for increasing hot-carrier stress. Device parameters as in Fig. 2. Characterization parameters: $f$ = 1 MHz, $t_{rise}$ = $t_{fall}$ = 100 ns, lower pulse level $V_L$ = -6 ... 0 V, amplitude $V_H$-$V_L$ = 5 V.

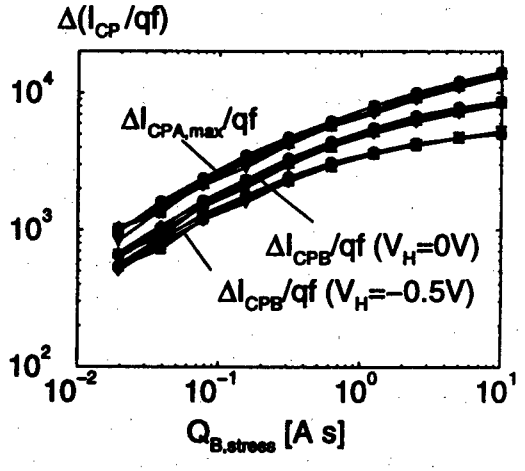



Fig. 4: $Q_{B,stress}$ dependence of the stress induced charge pumping currents $\Delta I_{CPA,max}$, $\Delta I_{CPB}(V_H$ = -0.5 V), and $\Delta I_{CPB}(V_H$ = 0 V) for $L_{drawn}$ = 1.0 µm, 1.2 µm, 1.5 µm, 2.0 µm, and 3.0 µm. W = 10 µm, further device and characterization parameters c.f. Figs. 2 and 3. Stress parameters: $M_{stress}$ = 8.45 %, $V_{D,stress}$ ($t_{stress}$ = 0, $L_{drawn}$ = 1 µm) = 7.6 V, $V_{G,stress}$ = 3.2 V. Characterization after $Q_{B,stress}$ = 10 As × $2^{k-10}$, $k$ = 1, 2, ..., 10.

Experimental CPA and CPB charge pumping results are shown in Fig. 4. Indeed, the charge pumping data show channel length independent damage. Thus, the parameters ($R_D$+$R_S$) and $\delta L$ after stress can now be determined using the above-presented method (c.f. Fig. 1).

## 4. Results

In Fig. 5 the stress induced degradations $\Delta I_{D,lin}/I_{D,lin}$, $\Delta R_D/R_D$ and $\Delta\delta L$ are depicted. Note, that in Fig. 5 b)

$$\Delta R_D/R_D = \Delta(R_D+R_S)/(1/2\times(R_D+R_S)) \quad (2)$$

holds, since only $R_D$ suffers from stress and $R_D=R_S$ is assumed before stress. For low values of $Q_{B,stress}$, $\Delta I_{D,lin}/I_{D,lin}$ and $\Delta R_D/R_D$ show similar $Q_{B,stress}$ dependencies proportional to $Q^n_{B,stress}$ with $n(\Delta I_{D,lin}/I_{D,lin}) \approx 0.21$ (for all channel length!) and $n(\Delta R_D/R_D) \approx 0.15$. For long stress times and large values of $Q_{B,stress}$, however, the drain series resistance degradation tends to saturate and even decrease, whereas the drain current continues to increase with a constant slope. Also the parameter $\Delta\delta L$ follows a relation $Q^n_{B,stress}$, but with $n(\delta L) \approx 0.5$, a value which is more than two times higher than the value of $n$ obtained for the other two parameters.

In order to interpret these findings we consider the schematic diagram in Fig. 6 explaining the different mechanisms taken into account. In the transition regime between channel and n⁻-region, the total conductivity is determined

1.) by a transistor field effect whose impact is a function of $V_G$ and the local threshold voltage,

and

2.) by the local resistance of the n⁻-regime.

In general, as the junction is approached from the channel, the contribution of 1.) is decreasing and the contribution of 2.) is increasing. After stress, negatively charged interface states (and perhaps in addition trapped electrons) in the damaged region lead to an increase of the local threshold voltage and to depletion of the surface of the n⁻-region within the drain. Thus the pn-junction shifts into the drain, i. e. the channel length indeed increases and the length of the n⁻-region decreases. However, in addition to this trend, the region deeper in the drain must be considered, where the total local conductivity is dominated by the contribution of the n⁻-resistance. There, a local stress-induced series resistance increase due to depletion of the region close to the oxide interface occurs. Thus, two mechanisms exist which have opposite impacts on $R_D$. Obviously, for short stress times the series resistance increase due to surface depletion dominates. For long stress times, however, the n⁻-region shortening effect compensates for the first effect and can achieve higher influence. The impact of these two effects leads to the behavior found in Fig. 5 b).

The parameter $\Delta\delta L$ (Fig. 5 c)) must be interpreted as an "equivalent channel length increase", since the measured data for this parameter are influenced by the following mechanisms:

- a local $V_{th}(y)$ increase (cf. Fig. 6d)),
- a local mobility decrease (not shown in Fig. 6 but possibly important for the whole damaged region), and
- a true effective channel length increase.

The decrease of $\Delta R_D/R_D$ at long stress times is a strong evidence that the length attributed to the n⁻-region decreases so that a true channel length increase indeed plays a role.

In the following, the consistency of the model is checked by an analytical evaluation. First, we calculate $\Delta I_{D,lin}/I_{D,lin}$ as a function of $\Delta R_D/R_D$ and $\Delta R_T/R_T$, where $R_T$ stands for the resistance of the inner transistor. Starting with

$$I_{D,lin} + \Delta I_{D,lin} = \quad (3)$$

$$\frac{V_D}{R_{tot,lin} + \Delta R_{tot,lin}} = \frac{V_D}{R_S + R_D + \Delta R_D + R_T + \Delta R_T},$$

where all Δ-terms = 0 before stress, we obtain:

$$\Delta I_{D,lin}/I_{D,lin} = \quad (4)$$

$$-\left[\frac{\Delta R_D}{R_D}\times\frac{R_D}{R_{tot,lin}+\Delta R_{tot,lin}} + \frac{\Delta R_T}{R_T}\times\frac{R_T}{R_{tot,lin}+\Delta R_{tot,lin}}\right]$$

Replacing $\Delta R_T/R_T$ in eq. (4) by $\Delta\delta L/L_{eff}$ we achieve the approximation

$$\Delta I_{D,lin}/I_{D,lin} \approx \quad (5)$$

$$-\left[\frac{\Delta R_D}{R_D}\times\frac{R_D}{R_{tot,lin}+\Delta R_{tot,lin}} + \frac{\Delta\delta L}{L_{eff}}\times\frac{R_T}{R_{tot,lin}+\Delta R_{tot,lin}}\right]$$

The drain current degradation predicted by this result is plotted in Fig. 7 together with the experimentally obtained data. In spite of the complicated $Q_{B,stress}$ dependencies of $\Delta R_D/R_D$ and $\Delta\delta L$ a simple power law is obtained for the calculated values of $\Delta I_{D,lin}/I_{D,lin}$ and an excellent agreement between measured data and model is achieved. Furthermore, the relative contribution α of $\Delta R_D/R_D$ to the total drain current degradation is depicted in this plot. As can be seen, for low stress times the degradation is dominated by the series resistance degradation. For long stress times, however, the contribution of the parameter $\Delta\delta L$ increases.

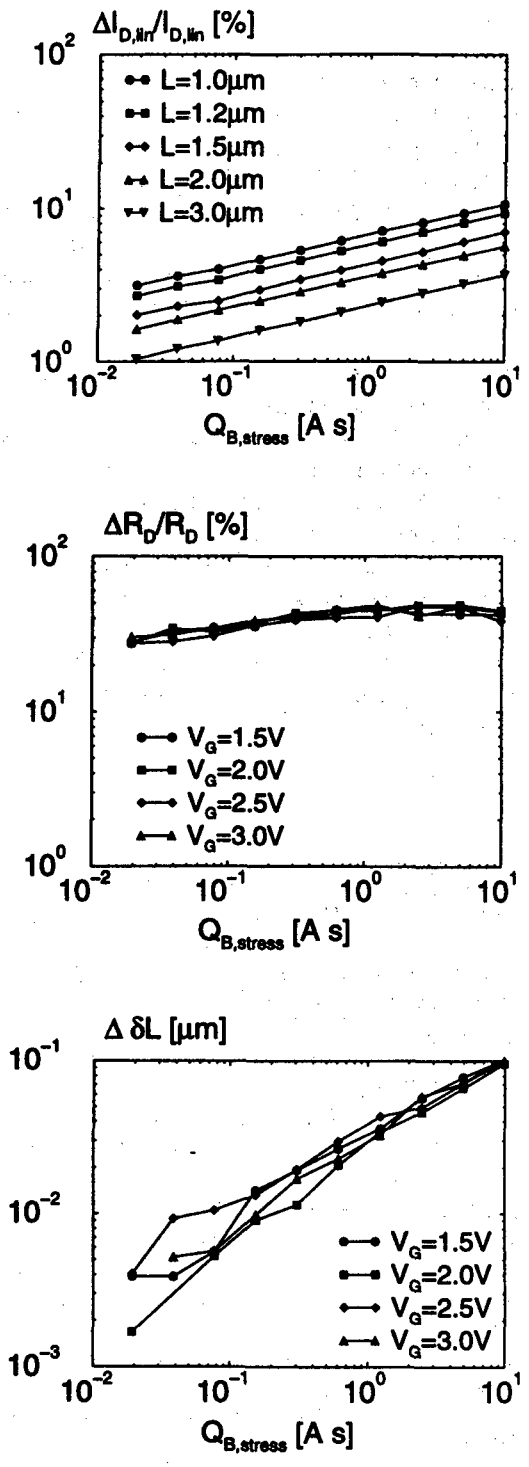



Fig. 5: $Q_{B,stress}$ dependence of the stress induced degradations $\Delta I_{D,lin}/I_{D,lin}$ (a), $\Delta R_D/R_D$ (b), and $\Delta \delta L$ (c). Device and stress parameters c.f. Fig. 2. Characterization drain voltage $V_D = 0.1$ V, Characterization gate voltage in a) $V_G = 3$ V.

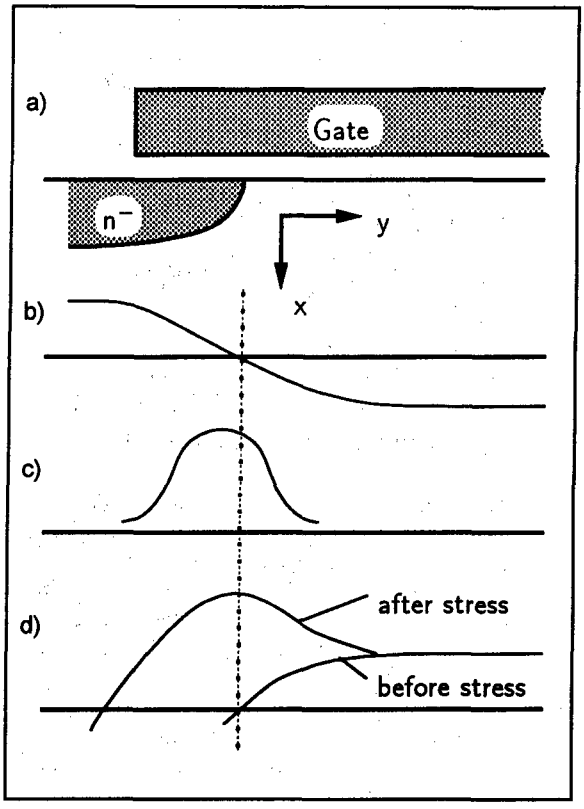



Fig. 6: Schematic diagram of the transition regime between channel and n⁻-region before and after stress. a) Cross section, b) effective doping, c) damage distribution, d) local $V_{th}(y)$ distribution.

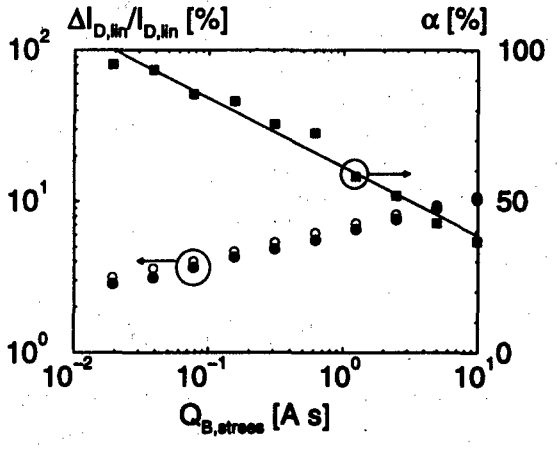



Fig. 7: Comparison of calculated (open symbols) and measured values (full symbols) for $\Delta I_{D,lin}/I_{D,lin}$ as a function of $Q_{B,stress}$. Furthermore, the relative contribution $\alpha$ of $\Delta R_D/R_D$ to the total drain current degradation is depicted.

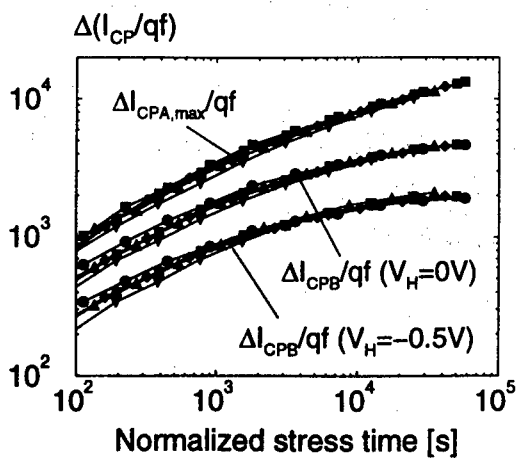



Fig. 8: Normalized stress time dependence of the stress induced charge pumping currents $\Delta I_{CPA,max}$, $\Delta I_{CPB}(V_H$ = -0.5 V), and $\Delta I_{CPB}(V_H$ = 0 V) for $L_{drawn}$ = 1.0 µm, 1.2 µm, 1.5 µm, 2.0 µm, and 3.0 µm after standard hot carrier stress. For further device and characterization parameters see Figs. 2 and 3. Stress parameters: $V_{D,stress}(L_{drawn}$ = 1 µm) = 7.8 V, $V_{G,stress}$ = 3.2 V ($I_{B,stress,max}$ condition), $M_{stress}(t_{stress}$ = 0, $L_{drawn}$ = 1 µm) = 8.45 %. Characterization after $t_{stress,norm}$ = 16 h × $2^{k-10}$, $k$ = 0, 1, 2, ..., 10. Normalization based on eq. (6) with $L_{drawn,0}$ = 1 µm.

Finally, we check the relevance of the data produced by our approach for standard stress experiments as used in most reliability investigations: A set of experiments was performed by adjusting $V_{D,stress}$ only at the beginning of the stress, so that $M_{stress}(t_{stress}=0) \neq f(L_{drawn})$ was obtained. For $t_{stress} > 0$, regulation was switched off, so that the electrical fields under stress, the amount and the distribution of damage etc. could develop in an uncontrolled way. In fact, it is found that after performing a channel length related stress time normalization procedure based on the definition

$$t_{stress,norm}(L_{drawn}) = t_{stress}(L_{drawn}) \times \frac{I_{B,stress}(L_{drawn})}{I_{B,stress}(L_{drawn,0})} \quad (6)$$

the experimentally obtained results for $\Delta I_{CPA,max}$, $\Delta I_{CPB}(V_H = 0$ V) and $\Delta I_{CPB}(V_H = -0.5$ V) coincide on single curves, respectively (Fig. 8). Although data with equal amounts and distributions of damage are not available here directly, the interpolated data reveal a similar behavior of device degradation as in our fully-controlled stress method.

## 5. Conclusion

A new method was presented which allows to quantitatively determine the drain series resistance degradation after hot-carrier stress. The impact of this effect and of an "equivalent channel length increase" contributing to the current degradation can be clearly distinguished. The new method is based on a stress technique which leads to equal amounts and distributions of hot-carrier induced damages in devices with different channel lengths. The damages are monitored by constant amplitude and constant base level charge pumping measurements. The relevance of the results obtained with the new method has been demonstrated by comparing these results with data from standard stress experiments.

Quantitative results have been shown for LATID-n-MOSFETs. For the investigated devices, series resistance degradation dominates the drain current degradation for short stress times whereas for long stress times the impact of the mechanisms attributed to the "equivalent channel length increase" prevails.

### References

[1] D. S. Ang et al., "A unified model for the self-limiting hot-carrier degradation in LDD n-MOSFET's", IEEE Trans. Electron Devices, 1998, p. 149

[2] Y. Pan et al., "Hot-carrier induced electron mobility and series resistance degradation in LDD n-MOSFETs", IEEE Electron Device Letters 1994, p. 499

[3] R. Thewes et al., "Mechanisms of hot-carrier degradation of analog device parameters in n-MOSFETs", Proc. ESSDERC 1993, p. 85

[4] B. J. Sheu et al., "Source and drain resistance of LDD MOSFET's", IEEE Electron Device Letters 1984, p. 365

[5] G. J. Hu et al., "Gate-voltage-dependent effective channel length and series resistance of LDD MOSFET's", IEEE Trans. Electron Devices 1987, p. 2469

[6] G. Groeseneken et al., "A reliable approach to charge-pumping measurements in MOS transistors", IEEE Trans. Electron Devices 1984, p. 42

[7] M. Tsuchiaki et al., "A new charge-pumping method for determining the spatial distribution of hot-carrier induced fixed charge in p-MOSFET's", IEEE Trans. Electron Devices 1993, p. 1768

[8] C. Chen et al., "Direct lateral profiling of both interface traps and oxide charge in thin gate MOSFET devices", IEEE Trans. Electron Devices 1998, p. 512

[9] P. Heremans et al., "Analysis of the charge-pumping technique and its application for evaluation of MOSFET degradation", IEEE Trans. Electron Devices 1989, p. 1318

PERGAMON

Microelectronics Reliability 38 (1998) 1069–1073

MICROELECTRONICS RELIABILITY

# On-wafer heating tests to study stability of silicon devices

D. Manic, Y. Haddab, R.S. Popovic

*Institute of Microsystems, Swiss Federal Institute of Technology – EPFL, CH-1015 Lausanne, Switzerland*

**Abstract**

A new on-wafer heating test structure is proposed for wafer level stability testing. The test structure is based on Wheatstone bridge configuration. Two out of four resistors are covered by aluminum. An optimization of the on-wafer heating technique is presented. The new structure and the structure consisting of two out of four resistors with the large openings in the silicon dioxide/nitride layers are measured. The piezoresitance effect was utilized to measure the mechanical stress in the silicon substrate. The mechanical stress is induced during the device fabrication. The effect of aluminum and silicon dioxide/nitride layers on the device stability was studied. A strong influence of aluminum layers on silicon device stability was observed.

© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

One of the most important goals in the design of silicon electron devices, and particularly silicon sensors, is stability. The silicon sensor stability may be seriously affected by long-term drift of mechanical stress induced during fabrication [1] and packaging. The drift of mechanical stress influences characteristics of sensor due to different physical phenomena such as piezoresistance effect [2], piezo-Hall effect [3] (for magnetic sensors) or dislocation generation and glide [4,5].

In this paper we present a continuation of our work in the field of wafer level stability tests [6], giving a comparison between two on-wafer heating test structures. We also give an analysis of the on-wafer heating procedure. The effects of metal and oxide/nitride layers on the electrical parameters of the test structures have been studied, as well as the permanent change induced by thermal stress.

## 2. Test structures and methods

The layouts of two test structures, 4R-MET and 4R-CO, accompanied by section views are presented in Fig.1 and Fig.2, respectively.

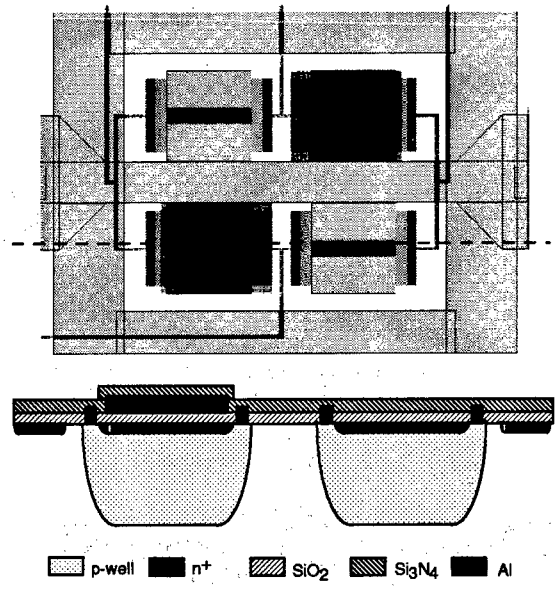



Fig.1. 4R-MET test structure, layout and section view.

The test structures are based on a Wheatstone bridge configuration. The bridge resistors were fabricated in the p-well and buried with the $n^+$ implantation. A heating ring, surrounding the

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00094-8

bridge, was made of the $n^+$ implantation. Technology details were given in [6] when the 4R-CO test has been introduced. In the test structure 4R-MET, two opposite out of four resistors are covered by the metal layer (Fig.1). The test structure 4R-CO consists of two out of four resistors with the large openings in the silicon dioxide and silicon nitride layers above them (Fig.2).

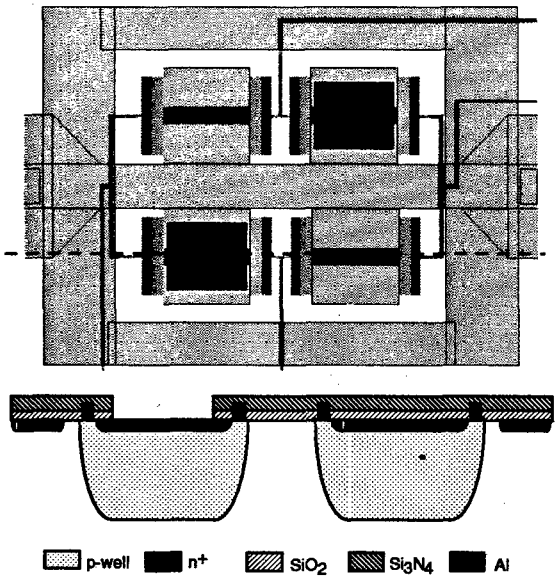



Fig.2. 4R-CO test structure, layout and section view.

We have applied heating power by biasing the low resistance ring around the test structure. The on-wafer heating system is presented in Fig.3.

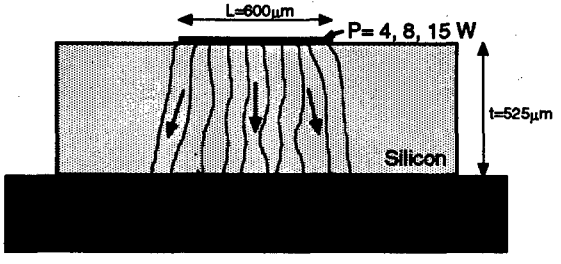



Fig.3. Schematic representation of the on-wafer heating system. Three heating powers (4,8 and 15W) have been applied during the experiments.

Assuming that the test structure is at uniform temperature when heating is applied, we roughly estimated the thermal resistance of the on-wafer heating system by

$$R_{th} = \frac{1}{\lambda_{Si}} \cdot \frac{t}{LW} = 13^{\circ}C/W, \quad (1)$$

where $\lambda_{si}$ is the thermal conductivity of silicon, $t$ is the thickness of the silicon wafer, $L$ and $W$ ($W$=400 $\mu m$) are the length and width of the test structure. The bridge input resistance was utilized as temperature sensor. A calibration of that resistance was performed on the packaged chip over the temperature range of 20°C to 300°C. During on-wafer heating, temperature has been increased for 200°C when the maximum power of 15W was applied. This was in agreement with the estimated temperature from the thermal resistance of the system, given by Eq.1. An important part of the on-wafer heating system was the thermostat (thermal chuck in probe station). There were two reasons for its utilization:

- to increase temperature during the on-wafer heating since $T_{test} = T_0 + R_{th} \cdot P$, where $T_0$ is the temperature of the thermostat and $P$ applied heating power ($T_0$=50°C).
- to decrease testing time since on wafer testing in industrial environment must be quick.

After the local on-wafer heating, the thermal chuck quickly stabilized temperature at the referent temperature level $T_0$ (temperature prior to the local on-wafer heating). The final heating procedure corresponded to 10s heating (longer heating produced the same effect on the 4R-MET Wheatstone bridge voltage) and 60s cooling-down (temperature stabilization).

As a result, the stability testing at the 4R-MET and the 4R-CO test structures was the following:

- offset measurement ($I_{bias}$=1mA, $T_0$=50°C)
- 10s heating
- 60s cooling down
- offset measurement ($T_0$=50°C)

We have applied three heating conditions with 4,8 and 15W heating power. The tests have been performed on both Float-zone (FZ) and Czochralski (CZ) wafers. The CZ wafers contain a concentration of oxygen of about $10^{18}cm^{-3}$, which is much higher compared to the FZ wafers (less than $10^{17}cm^{-3}$). This oxygen may influence propagation of the stress-induced dislocations, producing a

difference in the results of the accelerated test for the test structures realized in CZ or FZ type of wafers.

We also measured the temperature behavior of the offset using the thermal chuck system in the probe station (temperature range 25°C to 150°C) and local on-wafer heating (temperature range 25°C to 250°C).

## 3. Results and discussion

### *3.1. Offset voltage of the test structures and its stress interpretation*

Considering the piezoresistance effect as the origin of the voltage offsets for the given bridge configuration and biasing, the stress in the axisymmetric situation ($\sigma_x = \sigma_y = \sigma$, $\sigma_z << \sigma_x$, $\sigma_y$, xy-plane being the silicon surface) for p-type resistor in [110] crystallographic direction, is determined by

$$\sigma_1 - \sigma_2 = \frac{2 \cdot Voff}{R_0 \cdot I_{bias} \cdot (\pi_{11} + \pi_{12})}, \qquad (2)$$

where $\sigma_1$ is the stress at the surface of the resistor covered by the metal layer (4R-MET) or with the contact-opening (4R-CO), $\sigma_2$ is the stress at the surface of the normal p-well buried resistor, $V_{off}$ is the value of the offset voltage, $R_0$ is the input resistance of the bridge, $I_{bias}$ is the bridge bias current ($I_{bias}=1mA$), and $\pi_{11}$, $\pi_{12}$ are the piezoresistance coefficients of the p-type silicon ($\pi_{11}=6.6 \cdot 10^{-11}\ m^2/N$), $\pi_{12}=-1.1 \cdot 10^{-11}\ m^2/N$) [7].

#### *3.1.1. 4R-MET test structure*

The mean value of the measured 4R-MET offset was -630μV. The same results on the structures from two different fabrication runs have been obtained. The origin of the measured offset was the first point to be clarified. If voltage offset had a resistive origin, e.g. due to mask misalignment, the magnitude of the offset as a function of temperature would be linear with the same temperature coefficient as the bridge input resistance $\alpha_T$=0.7%/ °C. The measured 4R-MET voltage offset had a negative temperature coefficient $\alpha_T$=-0.1%/ °C in the range of 25 °C to 100 °C. The analysis of the measured temperature coefficient using known p-well resistance temperature coefficient and expected temperature coefficient of stress in the metallization layer [8], showed that the stress driven offset dominated the resistive one. Comparison between the temperature dependence of the measured voltage offset of the 4R-MET structure and calculated resistive voltage offset is given in Fig.4. This figure illustrates a non-resistive origin of the measured voltage offset.

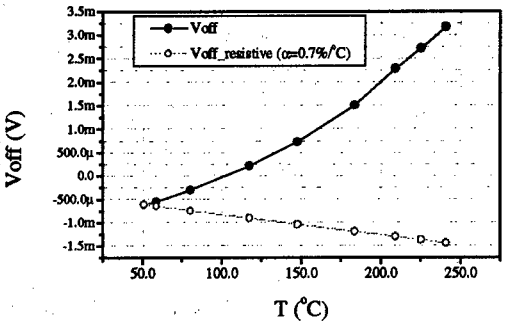



Fig.4. Comparison between the measured temperature dependence of 4R-MET voltage offset and the calculated resistive voltage offset temperature behavior.

Applying Eq.2, we calculated that the aluminum metallization above the resistors resulted in an additional mean compressive stress of –19 MPa at the silicon surface.

The aluminum film has been sputtered and then annealed. The annealing step leaves the film in a state of biaxial tension at room temperature. When the aluminum film is subjected to temperature cycle, due to greater thermal expansion coefficient of the aluminum film, tensile stress first relaxes and then goes into a state of compression [8]. Stress induced in the silicon substrate has the opposite sign (at the beginning compressive, then tensile). For a temperature higher than 200°C, a yield stress is reached and plastic deformation is occurring producing a stress hysteresis in the film and consequently in the silicon substrate. The calculated stress temperature hysteresis based on the measured offset temperature hysteresis (Eq.2) is presented in Fig.5.

Consequently, a difference of 7 MPa, compared to the stress level prior to heating, stayed after cooling down, corresponding to a drift in voltage offset of -230 μV.

Therefore, the electrical measurements on the 4R-MET test structure are very sensitive and moreover indicate a direct influence of the

deposited aluminum layer on the electrical parameters (such as resistance and offset) of the silicon devices fabricated below this aluminum layer.

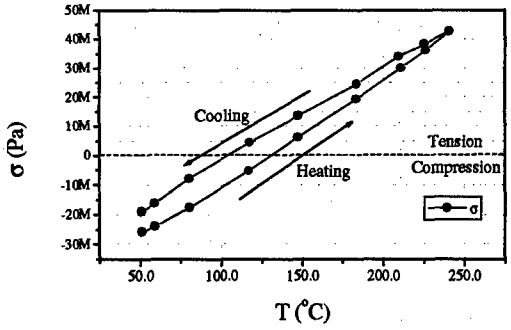



Fig.5. Stress vs. temperature plot for Si substrate under aluminum layer. The measurement has been performed on the 4R-MET test structure with the local heating.

*3.1.2. 4R-CO test structure*

The mean value of the measured offset was 5.25mV. The mean offset temperature coefficient was $\alpha_T$=0.8%/°C. Doing the analysis of the temperature coefficient for the 4R-CO test structure, we concluded that a resistive, and not mechanical stress induced voltage offset, was dominant. This was due to a fabrication problem, which increased the 4R-CO voltage offset. Although, this shift made it difficult to measure the absolute value of the voltage offset, for the stability tests only the relative changes in offset were important. In addition, we performed micro-Raman spectroscopy [9] on the 4R-CO structure. A tensile stress in the range of 150 to 250 MPa was measured for the silicon covered by silicon dioxide and nitride. The open silicon surface (silicon dioxide and nitride removed) was considered as referent, stress-free.

*3.2. Voltage offset change induced by thermal stress*

The results of change in offset after one on-wafer heating cycle are presented in Table 1.

The offset changes for both 4R-MET and 4R-CO test structures were observed. The induced thermal stress had a stronger effect on 4R-MET compared to 4R-CO since the mean change, for 15W heating power, were -215μV / -218μV (FZ/CZ), 116μV / 124μV (FZ/CZ), respectively. The influence of the aluminum layers on the device stability was strong, especially having in mind that the mean voltage offset of 4R-MET structure was only -630μV. In both structures we observed a quasi-permanent change because the change measured 24 hours after test stayed at 75% of the initial one.

Therefore, in the design of silicon devices, metal lines have to be placed as far as possible from the sensitive device zone, otherwise they may be a source of offset and offset instabilities with thermal drift. This practical conclusion, although known from the experience, has been more explicitly studied and presented here.

There was no significant difference between two wafer types (FZ and CZ) for the higher heating power. Standard deviation of the measured offset values was significantly higher in the case of Czochralski wafer. The offset change in CZ was lower for 4W heating power due to the pinning effect of oxygen interstitial atoms and $SiO_x$ clusters [6]. From the slight difference between FZ and CZ wafers, we concluded that stress-induced dislocation generation and glide is superimposed to the effect of the stress temperature hysteresis.

Table 1
Mean value and standard deviation (sd) of the measured voltage offset change of 4R-MET and 4R-CO test structures after on-wafer heating (15W,8W, 4W) on Float-zone (FZ) and Czochralski(CZ) wafer.

| Test structure | Heating power (W) | Mean ΔVoff (μV) **FZ** | sd ΔVoff (μV) **FZ** | Mean ΔVoff (μV) **CZ** | sd ΔVoff (μV) **CZ** |
|---|---|---|---|---|---|
| 4R-MET | 15 | -215 | 20 | -218 | 110 |
| 4R-MET | 8 | -152 | 11 | -134 | 54 |
| 4R-MET | 4 | -68 | 11 | -42 | 5 |
| 4R-CO | 15 | 116 | 23 | 124 | 163 |

## 4. Conclusions

A new on-wafer heating Wheatstone bridge test structure with two resistors covered by the aluminum layer has been proposed for thermally-induced stress wafer level stability testing. The optimization of the on-wafer heating technique has been presented with 10s heating and 60s cooling time. We reached 250°C by local on-wafer heating.

Mechanical stress at silicon surface covered by aluminum film, as well as the stress temperature hysteresis were measured. The measurements were based on the piezoresitance effect. The presented piezoresitance approach shows promising results.

Stability tests show that both the aluminum layer ($\Delta$Voff=-215$\mu$V) and the silicon dioxide/nitride layer ($\Delta$Voff=116$\mu$V) have an influence on test structure offset, although the influence of the aluminum layer is twice stronger. Thus, in the design of silicon devices, metal lines have to be placed as far as possible from the sensitive device zone, otherwise they may be a source of offset and offset instabilities with thermal drift.

Results of stability tests are slightly different for the FZ and CZ wafers showing an existing effect of oxygen concentration on stability. This effect is superimposed to the piezoresitance effect.

In conclusion, the on-wafer heating test structures are suitable for the measurements of the small electrical effects due to stress problems in semiconductor technology.

## Acknowledgments

The authors would like to thank Swiss National Science Foundation for funding this project.
The authors would also like to thank A.P. Friedrich for the useful discussions in the preparation of this paper.

## References

[1] S. M. Hu, "Stress-Related Problems in Silicon Technology," *Journal of Applied Physics*, vol. 70, pp. R53-R80, 1991.

[2] Y. Kanda and M. Migataka, "Effect of Mechanical Stress on the Offset Voltages of Hall Devices in Si IC," *Phys. stat. sol.*, vol. 35, pp. K115-K118, 1976.

[3] B. Halg, "Piezo-Hall coefficients of n-type silicon," *J. Appl. Phys.*, vol. 64, pp. 276-282, 1988.

[4] S. M. Hu, "Effects of surface films and film edges on dislocation movement in silicon," *Journal of Applied Physics*, vol. 46, pp. 1465, 1975.

[5] M. Reiche and W. Nitzsche, "The Influence of Stress on the Surface-Near Defect Structure," presented at Materials Research Society Symposium, 1992.

[6] Y. Haddab, D. Manic, and R. Popovic, "An On-Wafer Test Structure to Measure the Effect of Thermally-Induced Stress on Silicon Devices," *Microelectronics Reliability*, vol. 37, pp. 1441-1444, 1997.

[7] Y. Kanda, "A Graphical Representation of the Piezoresistance Coefficients in Silicon," *IEEE Transactions on Electron Devices*, vol. 29, pp. 64-70, 1982.

[8] W. Nix, "Mechanical Properties of Thin Films," *Metallurgical Transactions A*, vol. 20A, pp. 2217 - 2245, 1989.

[9] I. De Wolf, "Micro-Raman spectroscopy to study local mechanical stress in silicon integrated circuits," *Semiconductor Science and Technology*, vol. 11, pp. 139 - 154, 1996.

PERGAMON

Microelectronics Reliability 38 (1998) 1075–1080

MICROELECTRONICS RELIABILITY

# Characterization of SILC in thin-oxides by using MOSFET substrate current

B. De Salvo[a,b], G. Ghibaudo[a], G. Pananakakis[a], F. Mondon[b]

(a) LPCS/ENSERG, BP 257, 38016 Grenoble, France.

(b) CEA/LETI, 17 Avenue des Martyrs, 38054 Grenoble, France.

**Abstract**

A thorough SILC characterization by using MOSFET induced substrate current is for the first time developed. Based on obtained results, it is argued that a model consisting in a electrode-limited conduction, as the Fowler-Nordheim emission, cannot explains thin-oxide SILC results, while a bulk-limited trap-assisted transport well fits experimental data. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Stress-Induced Leakage Current (SILC) is commonly recognized as a very critical reliability issue of Flash non-volatile memories, limiting device miniaturization [1]. However, the physical origin and mechanisms of SILC have not yet been understood and several open questions remain.

In this work for the first time, a thorough SILC characterization by using MOSFET induced substrate current is presented. Based on obtained results, it is argued that a model consisting in a conduction limited by the Fowler-Nordheim emission cannot explains thin-oxide SILC results, while a bulk-limited trap-assisted transport well fits experimental data.

## 2. Experiment

Ginovker's carrier-separation method (see Figure 1-a) [2] has been used with metal-oxide-semiconductor field-effect transistors (MOSFETs) having as gate dielectric thermal $SiO_2$ with different thicknesses (40Å, 70Å, 100Å) and gate electrodes formed from n-degenerate polycrystalline silicon ($n^+$-poly).

In this experiment the gate of the n-channel MOSFET is biased at a high and positive voltage, then the gate current, due to electrons injected from the device channel into the oxide, and the "hole" substrate current are simultaneously recorded while varying $V_G$.

In order to quantitatively study SILC substrate current, we have performed constant current stress (CCS) between each consecutive recording of $I_G$-$V_G$ and $I_B$-$V_G$ characteristics.

The CCS condition for all cases was $J_G$=0.01 [A/cm²] (with positive gate bias), with an injected charge $N_{inj}$ varying between $10^{17}$ and $10^{20}$ [electrons/cm²].

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00085-7

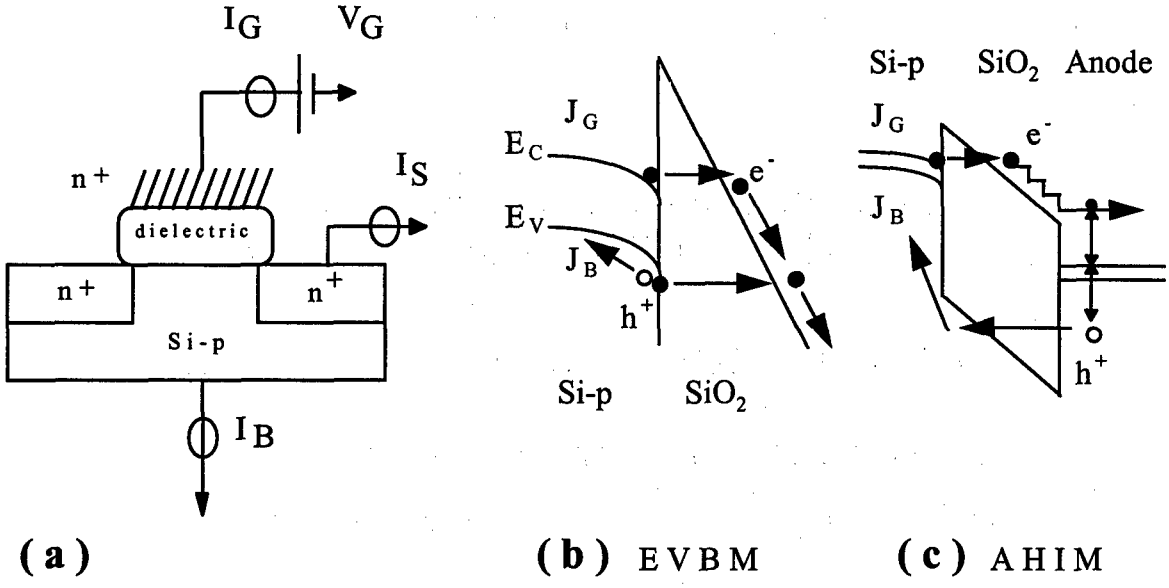



**Fig.1-(a)** Carrier-separation N-MOSFET.
**(b)** Electron Valence-Band Model (EVBM) [6]. **(c)** Anode Hole Injection Model (AHIM) [5].

## 4. Results and Discussion

In Figure 2 and Figure 3 measured gate current $J_G$, substrate current $J_B$ and quantum-yield $\gamma$, defined as:

$$\gamma = \frac{I_B}{I_G} \qquad (1)$$

as obtained on fresh samples and on stressed samples, are respectively shown.

The origin of the fresh substrate current has been extensively studied in our other works [3], [4]. We have demonstrated that, for thick $SiO_2$, the measured fresh substrate current is entirely due to Anode-Hole Injection (AHI) [5] (see Figure 1-c) while, for thin $SiO_2$, Electron Valence-Band (EVB) tunneling [6] (see Figure 1-b) plays the dominant role, especially at low electric fields.

In Figure 3-a it can be seen that the fresh quantum yield of the 40Å-thick $SiO_2$ is well fitted by the EVB Model while the 70Å and 100 Å ones are fitted by the AHI Model.

Figure 2-a clearly shows that stress-induced leakage current is easily detectable in all oxide gate current $J_G$, with the well-known thickness dependence [7] (namely the SILC decreases while increasing the oxide thickness).

On the contrary, concerning the substrate current $J_B$, the SILC appears significantly only in the thinnest oxide (see Figure 2-b).

For the 40Å-thick oxide the SILC components have been isolated from the pre-existing FN curve by subtracting the pre-stress gate current from the post-stress current, i.e.:

$$J_{SILC} = J_{STRESSED} - J_{FRESH} \qquad (2)$$

In Figure 3-b the 40Å-thick $SiO_2$ SILC quantum-yield:

$$\gamma_{SILC} = \frac{I_{B-SILC}}{I_{G-SILC}} \qquad (3)$$

appears to be quite invariant versus the oxide electric field, which means that $J_{G\text{-}SILC}$ and $J_{B\text{-}SILC}$ curves are rather parallel while varying the oxide field.

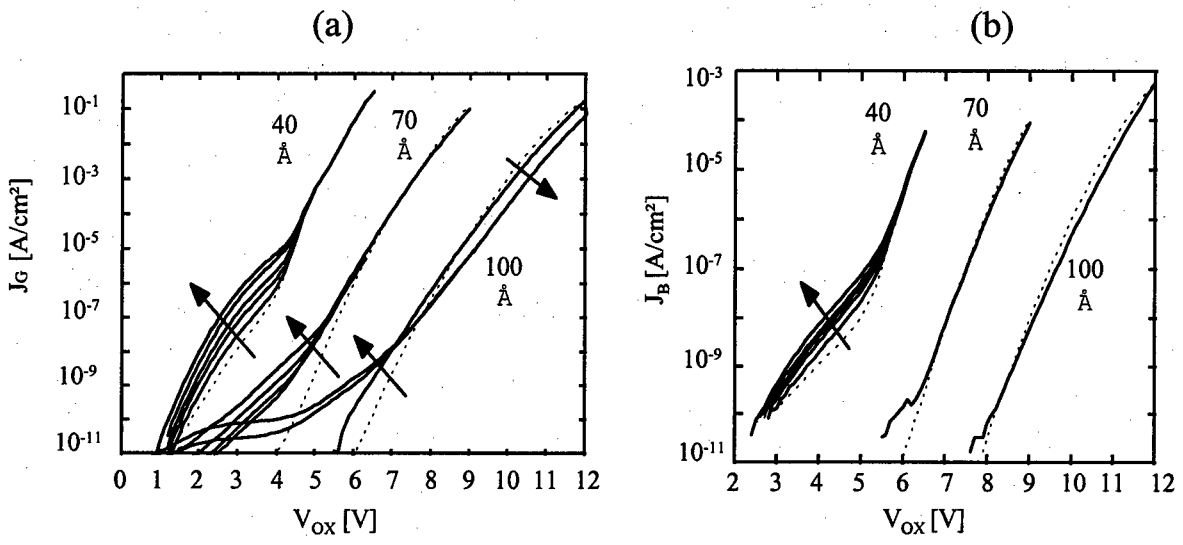



**Fig.2**- Gate current density $J_G$ **(a)** and substrate current density $J_B$ **(b)** [A/cm²] versus the oxide voltage $V_{OX}$ for the 40, 70, 100 Å-thick $SiO_2$ ,(dotted lines correspond to virgin samples, solid lines to stressed ones).

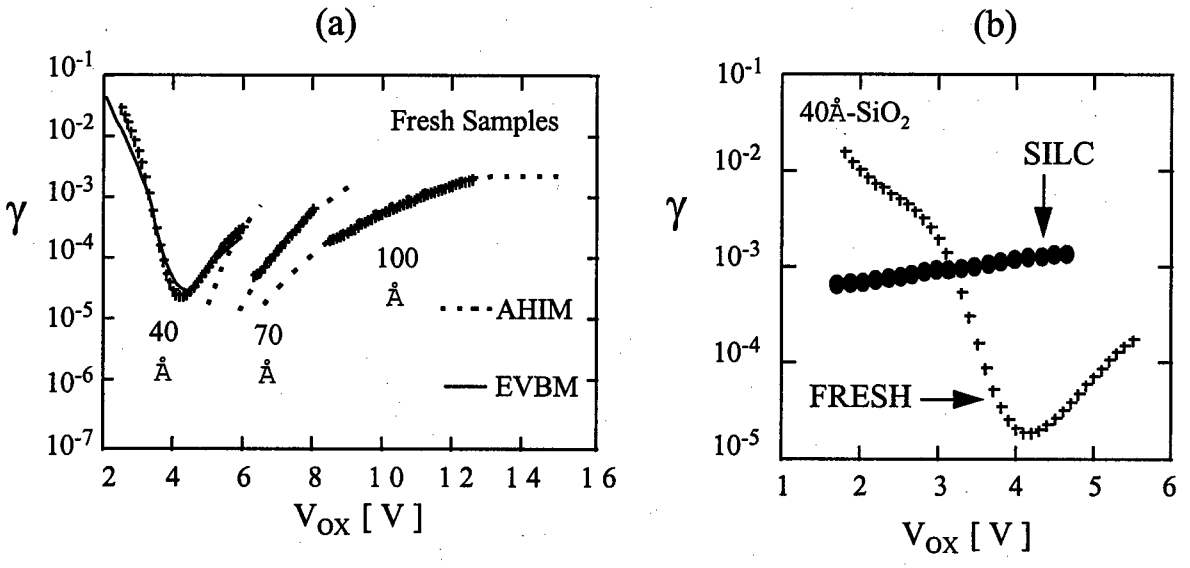



**Fig.3** - **(a)** Fresh quantum yield γ (=$I_B/I_G$) versus oxide voltage $V_{OX}$ for the 40, 70, 100 Å-thick $SiO_2$. Anode Hole Injection Model (AHIM) and Electron Valence Band Model (EVBM) are shown to fit quantum yield data (+).- **(b)** Fresh and SILC quantum yield $\gamma_{SILC}$ (=$I_{B\text{-}SILC}/I_{G\text{-}SILC}$) versus oxide voltage $V_{OX}$ for the 40 Å-thick $SiO_2$.

Moreover, Figure 4 shows that the 40Å-thick $SiO_2$ $J_{G\text{-}SILC}$ and $J_{B\text{-}SILC}$, at a fixed electric field value, are also parallel with the injected charge.

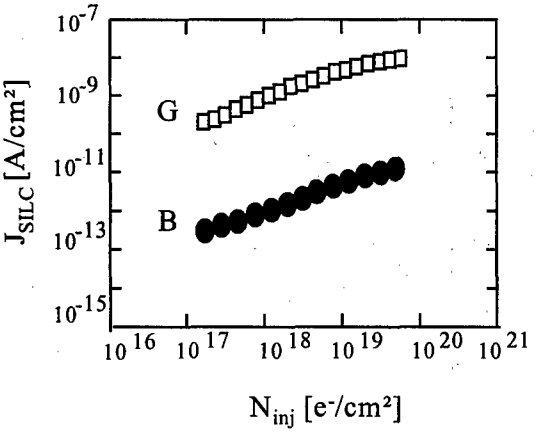



**Fig.4**- 40 Å-thick $SiO_2$ gate ( ) and substrate (●) SILC components ($J_{G\text{-}SILC}$, $J_{B\text{-}SILC}$ [A/cm²]), at a fixed gate voltage ($V_G$=4V), versus the charge injected during the stress ($N_{ini}$ [electrons/cm²]).

Figure 5 gives clear evidence for perfect equality of $J_{B\text{-}SILC}$ and $J_{G\text{-}SILC}$ kinetics with a characteristic power law, as expressed by [8]:

$$D_{SILC} = \frac{\delta \ln(J_{SILC})}{\delta N_{inj}} = K_n \cdot N_{inj}^{-\nu} \qquad (4)$$

indicating that the same defects are responsible for both bulk and gate SILC.

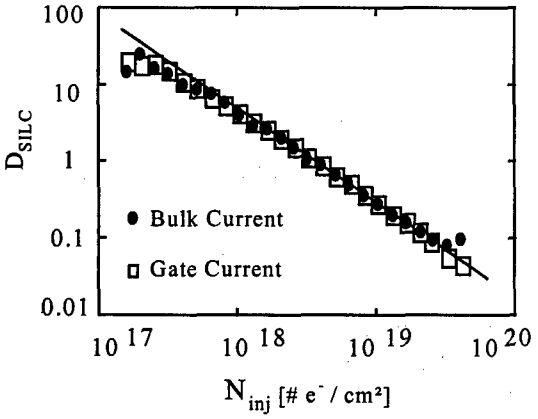



**Fig.5**- Gate and substrate SILC generation kinetics (40Å-thick $SiO_2$) versus the charge injected $N_{inj}$. Solid line corresponds to Eq.(4) with $K_n$=4.5·$10^{22}$ and ν=-1.22.

Stress-induced leakage tails in substrate currents was firstly observed by Schuegraf [5] for a 50Å device. Valence-band electron trap-assisted tunneling was invoked as a possible explanation. Based on our results, we conclude that SILC substrate current does consist in a valence-band electron trap-assisted tunneling detectable only in thin-oxides because thick-oxides post-stress substrate current is still dominated by the anode hole injection. Moreover we suggest that, after the stress, valence-band electrons are subjected to the same conduction mechanisms as conduction-band electrons, given that SILC substrate current behavior is identical to SILC gate current one (in terms of I-V characteristic slopes).

## 5. SILC Conduction Mechanisms

SILC conduction mechanism can be separated in two different phases (see Figure 6-a): (I) a trap filling-process by electron injection from the cathode; (II) an emptying-process, by electron trap-assisted transport in the oxide bulk. However, which of these two phases is the dominant mechanism remains even nowadays controversial.

In fact, Figure 6-b shows that if the analysis is limited to the gate current, then three main possibilities can be evoked to explain the SILC.

The first one consists in a Fowler-Nordheim (FN) tunneling from the cathode to a trap level in silicon dioxide (in Figure 6-b, a FN simulation with a barrier height $\Phi_{SC}$ of 1eV well fits $J_{G\text{-}SILC}$).

The second and third possibilities find consistence in the fact that a conduction law of the general form:

$$J = J_0 \cdot \exp\left(\frac{\beta}{k \cdot T} \cdot \sqrt{F}\right) \qquad (5)$$

with

$$\beta = \frac{1}{2} \cdot \sqrt{\frac{q}{\pi \cdot \varepsilon_0 \cdot \varepsilon_i}} \qquad (6)$$

again fits $J_G$ experimental data (see Figure 6-b), given $\varepsilon_i$=$\varepsilon_{SiO2}$=4. This law is normally invoked as describing a Schottky emission mechanism (given the factor ½ in β) or a Poole-Frenkel mechanisms under particular circumstances [9]. In fact, many criticisms have been advanced on the possibility of distinction between the Schottky and Poole-Frenkel mechanisms judging from the magnitude of a factor

in the exponent of the temperature-dependent term [9].

We argue that the analysis of the SILC substrate current allows us to eliminate the hyphotesis of an electrode-limited SILC conduction mechanism.

In fact, firstly, the parallelism of $J_{G\text{-}SILC}$ and $J_{B\text{-}SILC}$ rules out the FN tunneling mechanisms, where the slope of the current versus the electric field only depends on the barrier height. In Figure 6-b a FN simulation injecting electrons from the valence-band (with a barrier height $\Phi_{SV}$ of 2.1eV) does not fit $J_{B\text{-}SILC}$, at least at low and medium electric fields.

On the contrary, the Schottky or PF-like law well fits $J_{B\text{-}SILC}$ (see Figure 6-b), again given $\varepsilon_i=\varepsilon_{SiO2}=4$. Moreover, it should be mentioned that this temperature-activated law (5, 6) has been inferred after varying temperature. A SILC temperature analysis (limited to $J_G$ current), performed on the same 40 Å-thick $SiO_2$, has been extensively reported elsewhere [10].

In this case, using the substrate current results, the discrimination between Schottky and PF-like conductions comes from a physical evidence. We believe that a bulk-limited PF-like mechanism is the dominant phenomenon, finding no justification for an electron thermoionic emission from the valence-band of the silicon substrate.

Finally, the analysis of SILC substrate current shows that the dominant phenomenon in SILC conduction mechanisms is the PF-like emptying-process, at least at low and medium electric fields. At very high electric fields the FN injection filling-process may play a relevant role (see Figure 6-b).

## 6. Conclusions

For the first time at our knowledge, a quantitative SILC characterization by using MOSFET induced substrate current has been developed. Based on obtained results, it is argued that the dominant phenomenon in SILC conduction mechanisms is a bulk-limited trap-assisted emptying-process, at least at low and medium electric fields, while at very high electric fields the Fowler-Nordheim injection filling-process may play a relevant role.

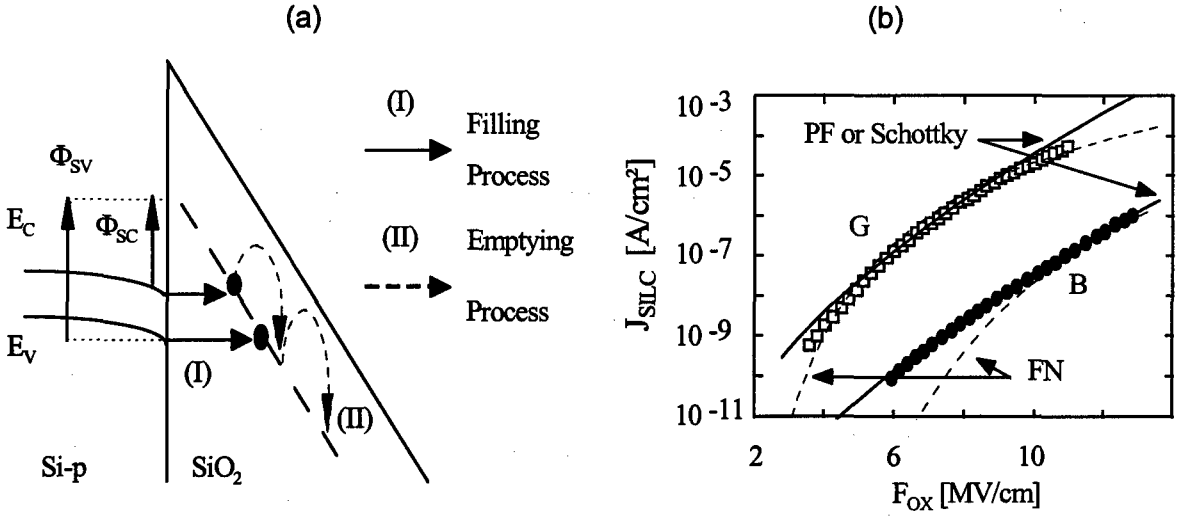



**Fig.6- (a)** SILC conduction mechanisms: (I) a trap filling-process by electron injection from the cathode ($\Phi_{SC}$, $\Phi_{SB}$ are FN barrier for conduction-band and valence-band electrons respectively); (II) an emptying-process by trap-assisted transport;
**(b)** Gate and Substrate SILC components ($J_{G\text{-}SILC}$, $J_{B\text{-}SILC}$) versus the oxide electric field. Solid lines correspond to Schottky or PF-like model ($\varepsilon_i=4$), dashed lines to FN model ($\Phi_{SC}$=1eV, $\Phi_{SB}$=2.1eV).

## References

[1] K.Naruke *et al.*, Proc. IEDM, p.424, 1988
[2] A.S.Ginovker *et al.*, Phys. Status Solidi, 26,1974.
[3] B.DeSalvo *et al.*, Int. Diel. Workshop, Toulouse (France), 1998.
[4] B.DeSalvo *et al.*, to be published on Solid State Electron.
[5] K.F.Schuegraf *et al.*, J. Appl. Phys., 76, 6, 1994.
[6] B.Eitan *et al.*, Appl. Phys. Lett., 43,1, 1983.
[7] E.Rosenbaum *et al.*, IEEE Tr. on El. Dev., 44, 2, 1997.
[8] A.Scarpa *et al.*, Microelectr. Engeen., 36,1997.
[9] J.J.O'Dweyer, The theory of electrical conduction and breakdown in solids dielectrics, Clarendon Press, Oxford, 1973.
[10] P.Riess *et al.*, 2nd French-Italian Symposium, $SiO_2$ and Advanced Dielectrics, Aquila (Italy), June 1998.

PERGAMON

Microelectronics Reliability 38 (1998) 1081–1084

MICROELECTRONICS RELIABILITY

# Hot Carrier Induced Device Degradation in RF-nMOSFET's

**Jong Tae Park[a], Byung Jin Lee[a], Dong Wook Kim[a], Chong Gun Yu[a], and Hyun Kyu Yu[b]**

*[a]Department of Electronics Eng., University of Inchon, #177 Dohwadong NamKu, Inchon, KOREA*
*[b]Semiconductor Div., ETRI, 161 Gajong, Yusung, Taejon, 305-350, KOREA*

**Abstract**

This study presents the hot carrier induced device degradation in 0.8 ㎛ RF-nMOSFET's within the general framework of the degradation mechanism. It has been found that the device degradation model of a single-finger gate MOSFET could be applied for a multi-finger gate RF-nMOSFET. The reduction of the cut-off frequency and maximum frequency after stress can be explained by the decrease of transconductance and the increase of drain output conductance. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

With the scaling towards an extreme submicrometer region, a cut-off frequency ($f_T$) of CMOS as high as 150 ㎓ has been reported[1]. Therefore, RF-CMOS is expected to replace the silicon bipolar transistors and GaAs MESFET's in RF front-end IC's for mobile telecommunocation devices in the near future. RF-CMOS is very attractive compared to the silicon bipolar transistor and GaAs MESFET's due to its lower manufacturing costs and lower power consumption. Moreover, the advantage offered by silicon technology is the ability to integrate RF circuits with other analog and logic circuits in a single chip for low cost. For the RF applications, some works about the device technology of RF-CMOS and the circuit performance of CMOS RF IC's have been published[2-5]. Because the hot carrier induced CMOS device degradation impacts on the performance of analog circuits, it is necessary to study the RF performance degradation of CMOS due to hot carrier effects. However, there is no study on the hot carrier induced device degradation for RF CMOS as far as we know.

This paper reports the hot carrier induced device degradation in 0.8 ㎛ RF nMOSFET within the general framework of the degradation mechanism.

## 2. Device fabrication and measurement

The nMOSFETs used in this work were fabricated using a 0.8 ㎛ twin well CMOS process on p-type silicon wafers with the resistivity of 2000 Ω·㎝. The gate oxide thickness is 175 Å, and a $TiSi_2$ silicide process was used to reduce the source-drain parasitic resistance. The gate pattern is multi-finger type with common source configuration as shown in Fig.1. All devices used in this work have 4 finger gates with unit finger width of 10 ㎛. The extracted gate resistance for 4 finger gates is about 80 Ω. Small signal scattering(S) parameters have been

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00111-5

measured on-wafer RF probes and with HP8510C Network Analyzer. To match the ground-signal probes of Cascade Microtech, special test structures have been designed with the MOSFET's common source-bulk configuration. On-wafer dummy structures were employed to de-embed pad parasitics. Measured S parameters were converted to admittance(Y) parameters, and the intrinsic device characteristics were calculated in the Y domain by using de-embedding technology. The parameters $f_T$ and $f_{max}$ have been determined as the frequency where the current gain is 0 dB and the frequency where the maximum available gain is 0 dB, respectively.

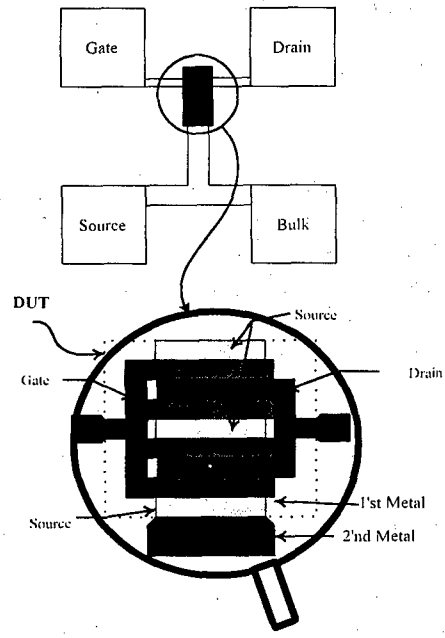



Fig.1. Test structure for RF-nMOSFETs

## 3. Device degradation model

Bulk MOSFET degradation can be attributed mostly to interface trap generation, although electron trapping in the gate oxide can also play a role. Since the main degradation mechanism due to hot carrier injection under the severest bias condition is interface trap generation, the device degradation such as drain current degradation, transconductance degradation, and threshold voltage shift has been modeled as follows[6] :

$$\Delta D = f\left[\frac{I_{ds}}{WH}\left[\frac{I_{sub}}{I_{ds}}\right]^{m} T_s\right] \quad (1)$$

where constant m and H are dependent on device processing technology and gate-drain bias voltage $V_{gd}$. $\Delta D$ has also a power law relation with stress time ( $\Delta D \propto T_s^n$ ). For bulk nMOSFET, it has been reported that the degradation parameter m is in the range of 2.5 to 5.5 and the degradation rate n is 0.3 to 0.5[7]. By using equation(1), we can extract the device lifetime and maximum allowable supply voltage, and also predict how munch the device degradation could be occurred at any bias condition just by monitoring the initial substrate current. To characterize the hot carrier induced device degradation of RF-nMOSFET, we applied the degradation model for bulk MOSFET to the muili-finger gate RF-nMOSFET.

Fig.2 shows the drain current degradation ( $\Delta I_{ds}/I_{ds}$ ) and the cut-off frequency degradation with the stress time for different gate bias voltages.

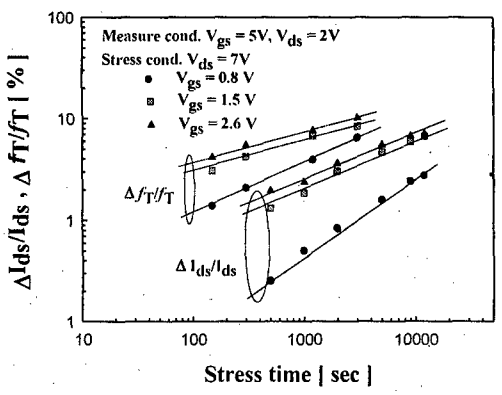



Fig.2. $\Delta I_{ds}/I_{ds}$ and $\Delta f_T/f_T$ versus stress time

The measurement condition was in the linear region. From Fig.2, the degradation rate n exhibits the stress bias dependence and lies in the range of 0.32 to 0.45. The RF performance degradation ( $\Delta f_T/f_T$ ) is severer than DC performance degradation ( $\Delta I_{ds}/I_{ds}$ ). To verify the applicability of the degradation model for RF-nMOSFET, the degradation parameter m has been extracted from the relation between $\tau I_{ds}$ and $I_{sub}/I_{ds}$ as shown in Fig.3.

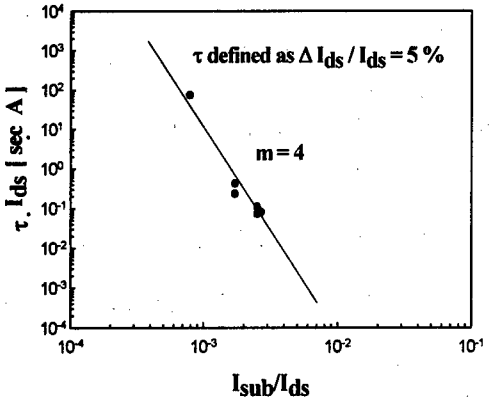



Fig.3. $\tau \cdot I_{ds}$ versus $I_{sub}/I_{ds}$ for RF-nMOSFETs

The extracted value of m is about 4, which is within the range of 2.5 to 5 for bulk MOSFETs. From Fig.2 and Fig.3, the device degradation model resulted from the hot carrier induced interface state generation can be applied for the lifetime prediction and hot carrier reliability of RF-nMOSFETs.

## 4. RF performance degradation

From the measurement of S parameters, it has been observed that, $f_T$ and $f_{max}$ of 0.8 ㎛ RF-nMOSFETs are, respectively, 12 ㎓ and 22 ㎓ at $V_{ds}$ = 5V and $V_{gs}$ = 3V. Since RF devices should have an $f_T$ that is about five to ten times higher than the operating frequency, our test devices can be used for the application of MMIC in the frequency range of 1~2 ㎓. The value of $f_T$ and $f_{max}$ for FET can be expressed as follows[8] :

$$f_T = \frac{g_m}{2\pi(C_{gd}+C_{gs})} \quad (2)$$

$$f_{max} = \frac{f_T}{2\sqrt{2\pi f_T R_g C_{gd} + g_{ds} R_{in}}} \quad (3)$$

where $C_{gs}$ and $C_{gd}$ are the gate to source and the gate to drain capacitance, $g_{ds}$ is the output conductance, $R_g$ is the gate resistance, and $R_{in}$ is the input resistance consisting of gate, source and channel components. Since $f_T$ and $f_{max}$ are function of $g_m$, $C_{gd}$, and $g_{ds}$ as can be seen in equation (2) and (3), RF performance could be degraded after DC hot carrier stress [9,10].

Fig.4 shows the measured S-parameters before and after stress. The bias condition for DC stress was Vgs =2V and Vds =7V. The S-parameters have been measured in the linear region ( Vgs =5V, Vds =2V ). The stress time was $T_S$ = 5000sec.

From Fig.4, $S_{21}$ and $S_{22}$ have been changed more significantly than $S_{11}$ and $S_{12}$ after stress. The degradation of $S_{21}$ and $S_{22}$ can be explained by the decrease of $g_m$ and the increase of $g_{ds}$ which results in the change of reflected parameters at output port.

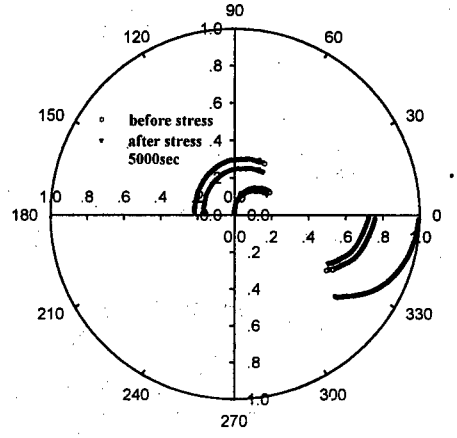



Fig.4. Measured S parameters before and after stress

Fig.5 shows the degradation of the current gain ($H_{21}$), maximum available gain (MAG), and maximum stable gain (MSG). The $H_{21}$ degradation is uniform with the frequency. However, MSG degradation is more significant than MAG degradation. The reason is that MSG is more sensitive to the degradation of $S_{21}$ than MAG

Fig.6 shows measured $f_T$ and $f_{max}$ as a function of drain current before and after stress. After stress, $f_T$ and $f_{max}$ are decreased and the degradation in the linear region is more significant than that in the saturation region. Since the delay-time in the logic gate is influenced by the degradation amount in the linear region and the performance of analog circuits is influenced by the degradation amount in the saturation region, the degradation of $f_T$ and $f_{max}$ will give a great impact on the CMOS RF IC's such as LNA, mixer and power amplifiers.

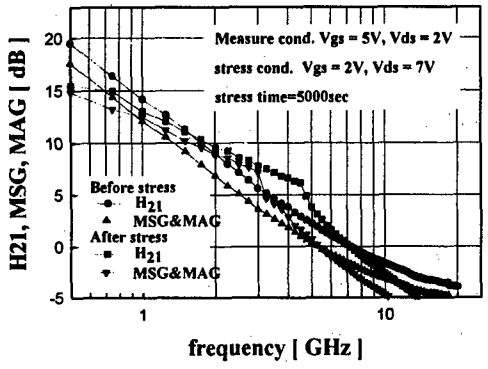



Fig.5. $H_{21}$, MSG and MAG versus frequency

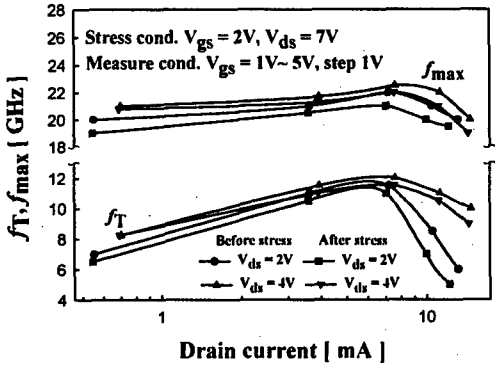



Fig.6. $f_T$ and $f_{max}$ versus drain current

## 5. Conclusion

Hot carrier induced device degradation of a single-finger gate MOSFETs can be used for mulit-finger RF-nMOSFRTs. After stress, $f_T$ and $f_{max}$ are decreased due to the hot carrier induced interface state generation. The degradation of $f_T$ and $f_{max}$ will pose serious constraints in RF IC reliability.

## Acknowledgement

This work was supported by Ministry of Education through the Interuniversity Semiconductor Research Center ( ISRC-97-E-1041 ).

## References

[1] H.S. Momose, E. Morifugi, T. Ohguro. M. saito. T. Morimoto, Y. Katsumata,H. Iwai " High-Frequency AC Characteristics of 1.5 ㎚ Gate Oxide MOSFET " Tech. Dig. IEDM-96, pp. 105-108, 1996.

[2] M. Saito, M.Ono, R. Fujimoto, H. Tanimoto, "0.15 ㎛ RF CMOS Technology Compatible with Logic CMOS for Low Voltage Operation", IEEE Trans. Electron Devices, vol. 45, pp. 737-736, 1998

[3] S.P. Voinigescu, S.W. Tarasewicz T. MacElwee. J. Ilowski " An Assessment of the State-of-the-Art 0.5 ㎛ Bulk CMOS Technology for RF Application " Tech. Dig. IEDM-95, pp. 721-724, 1995.

[4] A. Rofougaran, J. Y.-C. Chang, M. Rofougaran, A. A. Abidi " A 1 ㎓ CMOS RF Front-End IC for a Direct-Conversion Wireless Receiver " IEEE J. Solid-State Circuits. vol. 31, pp.880-889, 1996.

[5] T. Manku, "Microwave CMOS-Devices and Circuits ", Proc. of CICC, pp.59-66, 1998

[6] C. Hu, S.C. Tam, F.C. Hsu, P.K. Ko, T.Y. Chan, K.W. Terrill " Hot-Electron-Induced MOSFET Degradation-Model, Monitor, and Improvement IEEE Trans. Electron Devices, vol. 32, pp. 375-384, 1985.

[7] S.A. Kim, B. Menberu, T.E. Kopley, J.E. Chung " Oxide-Field Dependence of the NMOS Hot-Carrier Degradation Rate and Its Impact on AC-lifetime Prediction " Tech. Dig. IEDM-95, pp. 37-40, 1995.

[8] S. M. Sze, Physics of Semiconductor Devices, New York: Wiley, 1981, pp. 343-347.

[9] J. Chung, K.N. Quader, C.G. Sodini, P.K. Ko, C. Hu " The Effect of Hot-Eelectron Degradation on Analog MOSFET Performance " Tech. Dig. IEDM-90, pp. 553-556, 1990.

[10] Y.T. Yeow, C.H. Ling and L.KAh, "Observation of MOSFET degradation due to electrical stressing through gate-to-source and gate-to-drain capacitance measurement ", IEEE Electron Device Lett., vol. 21, pp. 366-368, 1991

PERGAMON

Microelectronics Reliability 38 (1998) 1085–1089

MICROELECTRONICS RELIABILITY

# The effect of hot electron current density on nMOSFET reliability

O.Buiu [a], S.Taylor [a], I.S. Al-Kofahi [b] and C.Beech [c]

[a] *Department of Electrical Engineering and Electronics, University of Liverpool, Brownlow Hill, Liverpool L69 3GJ, United Kingdom*

[b] *Hijawi College of Applied Engineering, Yarmouk University, Irbid, Jordan*

[c] *MITEL Semiconductors (formerly GPS) Tamerton Rd., Roborough, Plymouth, PL6 7BQ, UK*

## Abstract

In this paper, we consider the reliability of n-channel MOSFETs using the Substrate Hot Electron (SHE) technique. We confirm that there is a dependence of oxide degradation upon the current density during SHE injection (as previously observed by ourselves and others). In order to explain this effect, the detrapping of previously trapped electrons must be taken into account A new theoretical model is presented which accounts for the main features of the phenomenon. We consider the technologically important low field case (< 2MVcm$^{-1}$) for a range current densities (from 0.05 to 2 mAcm$^{-2}$ ) and injected charge densities up to 10 C/cm$^2$. The device lifetime for these different conditions is calculated and shown to be also a function of the current density. It is clear that in order to calculate the lifetime during normal operation from accelerated testing, the precise hot electron injection current density must be known, furthermore it must be demonstrated that the same degradation mechanisms hold at very high fields and/or current densities. This result has profound implications for device reliability predictions made using accelerated hot electron measurements and calls into question lifetime predictions made where the effect is not taken into account. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

The stability of the gate oxide properties tested under different conditions of charge injection has received much attention by various researchers because of the associated device degradation and eventual failure [1,2].

We consider here the reliability of n-channel MOSFETs using the Substrate Hot Electron (SHE) technique. We confirm that there is a dependence of oxide degradation upon the current density during SHE injection. The effect has been previously observed by ourselves and others [3,4,5]. However recent results show that in order to explain this effect the detrapping of previously trapped electrons must be taken into account and a new theoretical model is presented which accounts for the main features of the phenomenon. We also consider the technologically important low field case (< 2MVcm$^{-1}$) for a range current densities (from 0.05 to 2 mAcm$^{-2}$ ) and injected charge densities up to 10 C/cm$^2$.

This result has profound implications for device reliability predictions made using accelerated hot electron measurements which we therefore explore in this paper.

## 2. Experimental set-up

The n-MOSFETs investigated had a n$^+$-poly–Si

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00074-2

gate with a dry $O_2$ oxide of thickness 15nm fabricated using a 0.7μm CMOS process.

The hot electrons were provided by an underlying forward biased p-well/ n-substrate junction. The hot electron energy was controlled by varying the p-well bias ($V_w$)The average field in the oxide ($E_{ox}$) was determined by the gate voltage ($V_g/d_{ox}$). The injection current ($J_{inj}$)was provided by varying the substrate voltage according to [4].

Extensive 2D simulations were carried out using the simulator MEDICI in order to ensure uniform injection conditions. The oxide charge was measured by the shift in subthreshold transfer characteristics, the interface state density was measured using 2 and 3 level charge pumping [6].

## 3. Experimental results and discussion

Fig.1 (symbols) shows the increase in trapped charge ($\Delta N_{ot}$) under different $J_{inj}$. An increase in the trapped charges with increasing $J_{inj}$ is seen. These injected electrons can free trapped electrons and cause detrapping during injection.

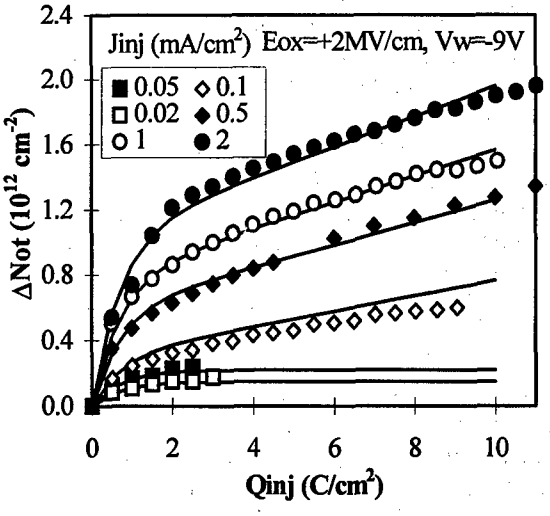



Fig. 1 Dependence of the trapped charge on the injected charge, for different current densities

To see the effect of electron injection on the electron detrapping, two devices were stressed by SHE under +2MV/cm and $J_{inj}=5x10^{-4}A/cm^2$. Electrons trapped during the stress are shown in fig.2 (symbol ' Δ'). After injecting 3 C/cm$^2$, in one device, $J_{inj}$ was reduced to zero (symbol ' • '), while for the other $J_{inj}$ was reduced to $5x10^{-5}$ A/cm$^2$ (symbol 'o').

As shown in fig.2, negligible change in the trapped electron density was observed when electrons were not injected, and this result confirms earlier finding [3].

However if, following injection at $5x10^{-4}$ A/cm$^2$, electrons continue to be injected at $5x10^{-5}$ A/cm$^{-2}$, then significant detrapping takes place confirming that the presence of hot electrons in the oxide is necessary for electron detrapping to occur.

This clearly points to same form of impact ionisation being responsible for the detrapping, as proposed originally by Nissan-Cohen et al [1].

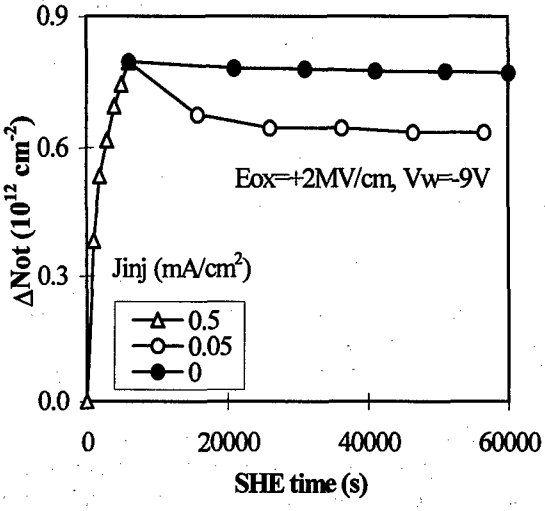



Fig. 2 The effect of hot electron injection upon electron detrapping

In a second experiment electrons were injected at $5x10^{-4}$ A/cm$^2$ and the net trapped charge was monitored at $Q_{inj}$=0.5 C/cm$^2$ intervals. On an identical device electrons were injected at $5x10^{-4}$ A/cm$^2$ for 0.45 C/cm$^2$, then $5x10^{-5}$ A/cm$^{-2}$ for 0.05 C/cm$^2$, making the time for injection roughly twice as long as in the first case.

During this period of reduced electron current density, the net charge level was significantly reduced (Fig. 3), confirming again the current density dependence of detrapping.

However, this current density dependence of trapped charges was not observed during substrate hot hole injection [7], as can be seen in Fig. 4. This is probably due to the fact that holes are much less mobile in $SiO_2$ than electrons hence collision events

of injected holes with trapped holes are much more rare than in the electronic case.

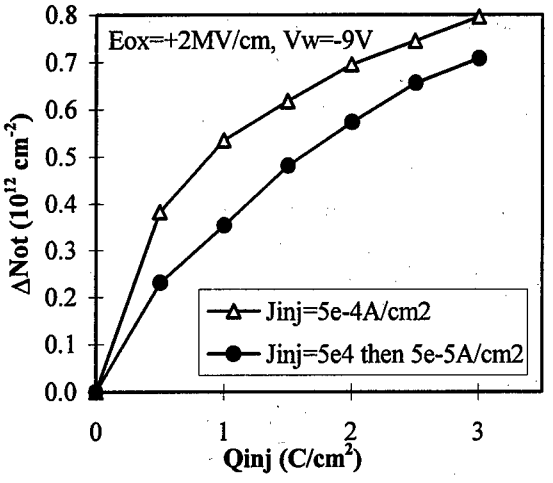



Fig. 3 Some of the electrons trapped during high $J_{inj}$ are detrapped during the low $J_{inj}$ period

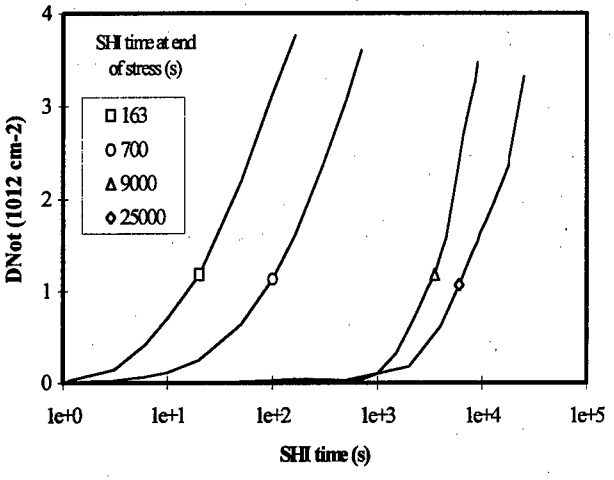



Fig. 4 For the same dose of injected holes under different injection current density, the same amount of traps were trapped.

### *3.1 Physical model and parameters extraction*

The above results suggest that the first order trapping model with single capture cross section [8, 9] must be modified to include the detrapping of trapped charges during injection. In cases were there might be both the filling of existing bulk traps, detrapping and the generation of new charged defects, a general equation under the assumption that mechanisms are independent of each other is:

$$n_M(t) = \alpha J_{inj}^{\beta} N_o [1 - \exp(-J_{inj}\, t\sigma / q)] + \gamma J_{inj}^{\eta} t \quad (1)$$

where:

- $n_M(t)$: measured trapped charge,
- $N_0$: concentration of as grown (native) traps with a capture cross section $\sigma$,
- $\alpha$ and $\beta$: occupation function parameters;
- $\gamma$ and $\eta$: generation coefficients.

A power law has been used to describe interface state and bulk charge build up during channel hot carrier injection at the drain [10], and this approach was followed here also, in defining the generated trap term.

Using (1) to fit the experimental data (using a least squared errors method), we obtained the following values for the parameters:

- $N_0=1.1\times10^{12}cm^{-2}$, $\sigma=2.2\times10^{-19}cm^2$,
- $\alpha=12.25$, $\beta=0.4176$,
- $\eta=1.23$, and $\gamma=0.362$ for $J_{inj}>5\times10^{-5}A/cm^2$ (0 for $J_{inj}\leq5\times10^{-5}A/cm^2$).

The trapped charges calculated from (1) are shown in Fig.1 (lines). There is a good agreement between the experimental and calculated results over the whole range considered.

The disappearance of any injection current dependence for the high oxide field (>4 MV $cm^{-1}$) is likely to be due to trap generation effects which are known to increase at high fields due to increased energy of electrons in the oxide [2]. If the trap generation rate greatly exceeds the detrapping rate then the dynamic balance is disturbed in favour of generation and the effect of detrapping during stress much reduced.

Figure 5 shows the occupation and the generation functions as functions of $J_{inj}$. Clearly negligible generation occurs at low $J_{inj}$ and both generation and trap occupation increase with $J_{inj}$.

Although the precise nature of the microscopic defects responsible for the degradation are unknown, a qualitative understanding of the behaviour presented in figures 1-3 can be gained by applying the model for detrapping kinetics presented in [11]. The model was previously used to describe the electron detrapping in combined avalanche electron injection (AEI) and Fowler-Nordheim (FN) stress experiments on MOS structures.

It is generally agreed that traps with capture cross-sections in the range $10^{-18}$ - $10^{-19}$ are water or hydrogen related (see [2] and ref. therein).

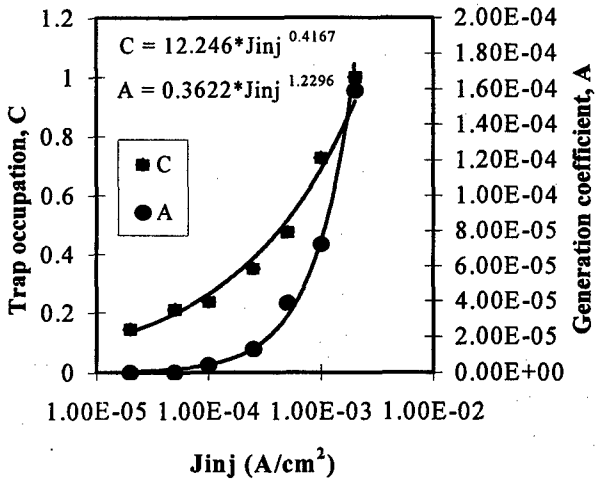



Fig. 5 Trap occupation ($C = \alpha J_{inj}^{\beta}$) and trap generation ($A = \gamma J_{inj}^{\eta}$) during SHE as a function of $J_{inj}$.

Representing these native traps by ≡Si - H, the following scheme emerges:

$$\equiv Si - H + e^- \rightarrow (\equiv Si-)^- + H \quad (2)$$

$$(\equiv Si-)^- + X \rightarrow \equiv Si\bullet + Y \quad (3)$$

$$\equiv Si\bullet + e^- \rightarrow (\equiv Si-)^- \quad (4)$$

In the first part of the curve presented in fig. 2, we can see an important increase in the trapped charge. The injected charge is trapped on both already existing traps ('as grown' or native traps, like ≡Si - H in Eq. 2) and traps created during the injection. Even detrapping is present, the balance of the whole process is in favour of trapping electrons (reaction described by Eq. (2)).

If now the current density is reduced to zero, both trapping and detrapping cease. If electron injection is resumed but at a lower $J_{inj}$ the dynamic balance is altered in favour of detrapping due to the presence of hot electrons and the involved time period (Eq. (3)). A further passivation reaction may also occur due to the reaction of liberated hydrogen with the active ≡Si •.

$$\equiv Si\bullet + H \longrightarrow \equiv Si-H \quad (5)$$

generating a decrease in the trapped charge.

*3. 2 Reliability implications*

We now consider the reliability implications of the above. Based on the experimental data ($\Delta N_{ot}$ vs. $Q_{inj}$, for different current densities), we calculated the shift in threshold voltage of the devices as a function of injection current density, at a given number of injected electrons per unit area (see Fig. 6).

Clearly the amount of shift depends upon a power law function of the current density, for a given no of injected electrons.

We further investigate the effect of different current densities upon the device 'lifetime', defined here as the necessary time for a device to reach a specified shift in $V_t$. The evaluation procedure was as used in [5], where it was applied for insulated

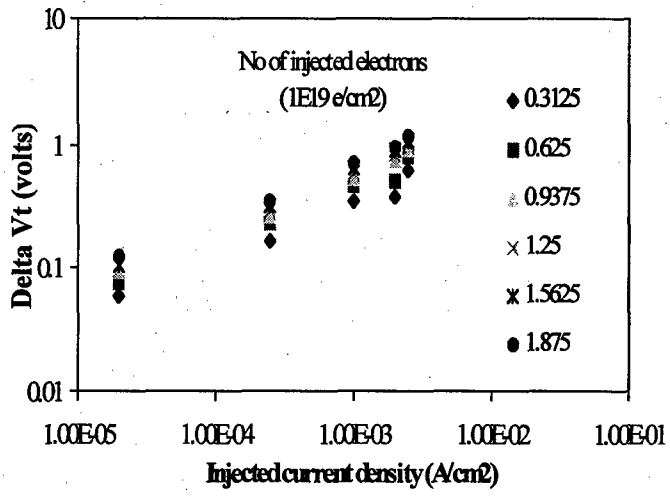



Fig. 6 Shift in the $V_t$ voltage as a function of injection current density, at a given number of injected electrons

gate field effect transistors (IGFET's) subject to electron stressing by a pulse injection technique (PIT).

Using the parameters from a least square fitting of data presented in Fig. 6, we calculated $N_{inj}$ (for a given $\Delta V_t$), device lifetime - $t_l$ was then given by:

$$t_L = \frac{qN_{inj}}{J_{inj}} \quad (6)$$

Results of these calculations - for ΔVt equal to 10, 50, and 100 mV, are presented in Fig. 7.

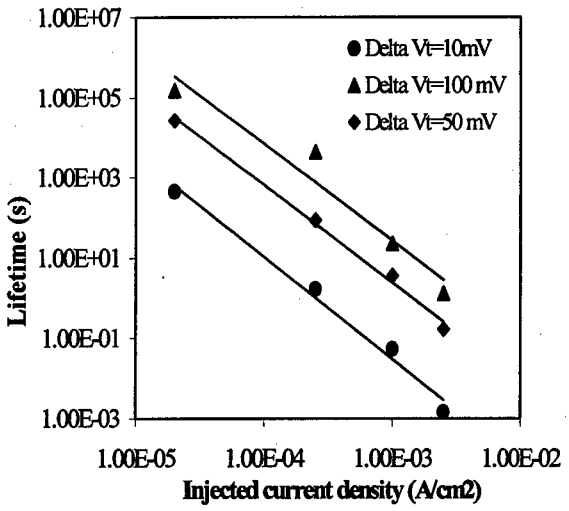



Fig. 7 Calculated lifetime vs. $J_{inj}$, for three different specified threshold voltage shift

It is clear that in order to calculate the lifetime during normal operation from accelerated testing, the precise hot electron injection current density must be known, furthermore it must be demonstrated that the same degradation mechanisms hold at very high and/or current densities.

## 3. Conclusions

Present paper confirms the existence of a dependence of oxide degradation upon the current density during SHE injection, observed here for the low field case ($\leq 2$ MVcm$^{-1}$).

We are presenting a new theoretical model describing this effect, taking into account the detrapping of previously trapped electrons. The model shows good agreement with the experimental results over the whole range considered. More than that, we are also suggesting a possible reaction mechanism which might be behind the observed results. The effect is important also for reliability predictions - using accelerated tests - emphasising the importance of an accurate determination of hot electron injection current density and calls into question lifetime predictions made where the effect is not taken into consideration.

## Acknowledgements

The authors would like to acknowledge the EPSRC for the provision of funding, the Department of Electrical Engineering and Electronics (University of Liverpool) for the provision of facilities, and finally MITEL Semiconductors (formerly GPS) for the test wafers. .

## References

[1] Y.Nissan-Cohen, J.Shappir, and D.Frohman-Bentchkowsky J. Appl. Phys. 60 (1986), 2252-2024.
[2] D.J.DiMaria and J.Stasiak J. Appl. Phys. 65 (1989), 2342-2356.
[3] A.v.Schwerin and M.M.Heyns, in: W.Eccleston and M.Uren (Eds.), Insulating Films on Semiconductors (Proc. INFOS 91), Adam Hilger, Bristol , 1991, pp. 263 - 266.
[4] S.P.Zhao,S.Taylor and A.McPhie, Microelectronics Journal 25, 515, 1994.
[5] H,S.Kim, A. Reisman and C.K.Williams, J. Electrochem. Soc. 144 (7), 2517-, 1997
[6] M.J.Kivi and S.Taylor, Microelectronics Journal, 27, 687 - 2521, 1996.
[7] I.S. Al-Kofahi, The Hot Carrier Induced Degradation of $Si/SiO_2$ Interface, PhD Thesis, Liverpool "John Moores" University, UK, 1997.
[8] T.H. Ning, Solid State Electron. 21 (1978), 273 - 282.
[9] D.J. DiMaria and J.W. Stasiak, J.Appl.Phys. 65 (1989), 2342 - 2356.
[10] B.Doyle, M.Bourcerie, J.C.Marchetaux and A.Boudou, IEEE Trans on Electron Dev. 37 (1990), 744 - 754.
[11] J. F. Zhang, S. Taylor, and W. Eccleston, J. Appl. Phys. 71 (1992), 5989 - 5996.

PERGAMON

Microelectronics Reliability 38 (1998) 1091–1096

MICROELECTRONICS RELIABILITY

# Dependence of gate oxide breakdown on initial charge trapping under Fowler-Nordheim injection

A. Martin[a,b], R. Duane[a], P. O'Sullivan[a], A. Mathewson[a]

[a] Technology Characterisation and Modelling Group (TCMG),
National Microelectronics Research Centre (NMRC),
Lee Maltings, Prospect Row, Cork, Ireland,
Tel: + 353-21-904201, Fax: + 353-21-270271,
e-mail: rduane@nmrc.ucc.ie, Web-page: http://nmrc.ucc.ie/

[b] on leave to:
Siemens AG, HL ZUV MON,
Wafer Level Reliability Engineering,
Otto-Hahn-Ring 6, 81739 München, Germany,
Tel.: + 49-89-636-45257, Fax: + 49-89-636-45822,
e-mail: andreas.martin@hlistc.siemens.de, Web-page: http://www.siemens.de/semiconductor/

**Abstract**

This work demonstrates that for constant oxide reliability stresses in the Fowler-Nordheim regime a low initial rate of charge trapping/detrapping results in long times to breakdown. It was found for MOS gate oxides that when the initial trapping has been completed at low fields times to breakdown enhance. Depending on the stress sequence measurement results can vary significantly which is of great relevance for correct oxide lifetime predictions. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Constant stress measurements: Constant Voltage Stress (CVS) and Constant Current Stress (CCS) are commonly performed for oxide reliability assessment (see [1] and its references). In order to keep the measurement times reasonably short these stresses are usually highly accelerated at fields in the Fowler-Nordheim (F-N) tunneling regime where charge trapping/detrapping and trap generation occur in the oxide. It has been shown [2,3] that the quantity of trapped charges and generated traps vary depending upon the stress field.

The aim of this work is to demonstrate the impact of charge trapping under F-N injection on the time ($t_{bd}$) and the injected charge ($Q_{bd}$) to breakdown. This is performed by investigating the trapping characteristics on the basis of current-time (I-t) curves of CVS or voltage-time (V-t) curves of CCS under F-N injection. Capacitance-voltage(C-V) or current-voltage (I-V) measurements at low

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00138-3

fields during the constant stress measurements have not been employed in this work because they could affect the trapping characteristics [4–8].

It is shown here that the rate of the initial trapping/detrapping during CVS and CCS plays an important role for oxide degradation. The relevance of this study lies in the fact that $t_{bd}$ and $Q_{bd}$ can change depending on the reliability stress sequence. As a result, oxide lifetimes at operating conditions and field acceleration parameters deviate from their true values.

## 2. Experimental details

All experiments were performed on n-type Si-substrate/$SiO_2$/poly-crystalline-Si MOS gate oxide test capacitors produced in an industrial process. Thermally grown oxides in a thickness range of 10-20nm with oxide areas of $10^{-6}$–0.16cm$^2$ were tested. The oxide test structures consisted of no LOCOS edges and had no poly-gate silicidation. The poly gates were degenerately doped and gate and substrate contacted via Al-pads on field oxide. Measurements were performed in accumulation with a positive gate potential at room temperature. The measurement samples consisted of 30 or more capacitors equally distributed across the 6 inch wafers. All measurements were performed with a HP 4062 Parametric Test System and a half-automatic Wentworth Wafer Prober.

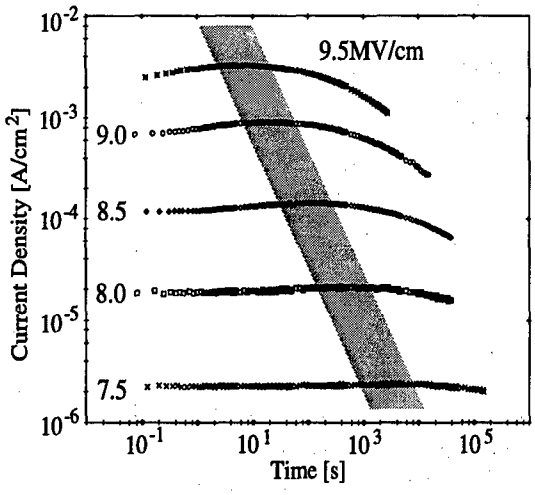



Figure 1: I-t curves of a 10.9nm gate oxide with $A_{ox}$=7.84$x$10$^{-2}$cm$^2$. Breakdowns are not displayed.

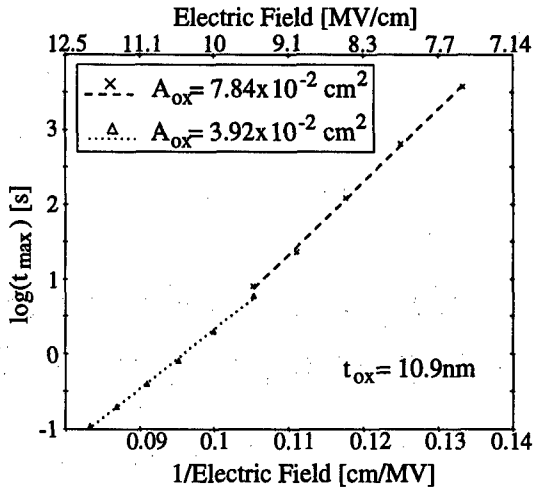



Figure 2: $t_{max}$ versus $1/E_{cvs}$ for a 10.9nm gate oxide.

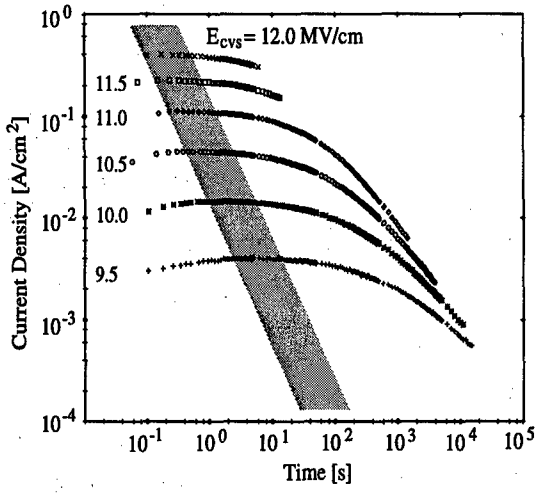



Figure 3: I-t curves of a 10.9nm gate oxide with $A_{ox}$=3.92$x$10$^{-2}$cm$^2$. Breakdowns are not displayed.

## 3. I-t characteristics

CVS measurements over a wide bias range from 6.5–12.0MV/cm were performed on a 10.9nm oxide. I-t curves of these measurements are presented in Fig. 1 for a field range of 7.5–9.5MV/cm. Typically, the current decreases for all I-t curves beyond the maximum current ($I_{max}$). A decrease in current is due to dominant electron trapping in the bulk of the oxide [2,4,9,10]. $I_{max}$ is illustrated by the grey area in Fig. 1 and all subsequent I-t characteristics.

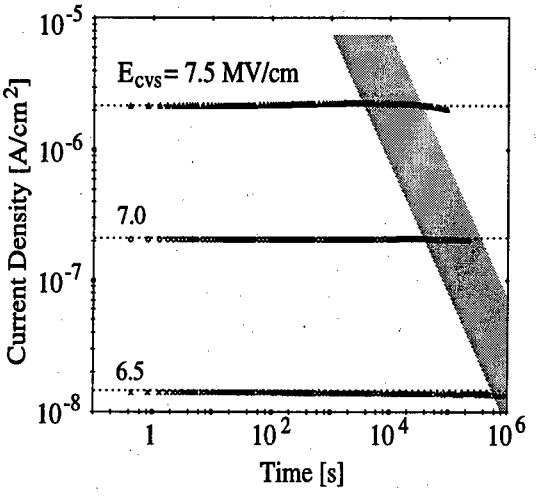



Figure 4: I-t curves of a 10.9nm gate oxide with $A_{ox}$=1.64$x10^{-1}$cm$^2$. Breakdowns are not displayed.

The corresponding time ($t_{max}$) of $I_{max}$ is strongly dependent on the stress field, as shown in Fig. 2. $t_{max}$ decreases with increasing field. For stress biases greater than 12MV/cm it was no longer possible to monitor $t_{max}$ and $I_{max}$ accurately because the time resolution of the measurement hardware was limited.

The initial current increase which can be clearly seen in Fig. 1 and 3 is believed to be the result of detrapping of native electrons [2–4] which were introduced during processing. The initial detrapping is a finite process which saturates at $I_{max}$. We have observed from experiments that the current increase vanishes for fields below 7.0MV/cm which is in agreement with the trap occupation model of Nissan-Cohen et.al.[2].

The I-t curves of Fig. 4 indicate that below 7.0MV/cm the initial current increase is no longer existent for the 10.9nm oxide under investigation. The horizontal dotted lines can be used as guidelines to recognise a current decrease/increase. Clearly, the I-t curve of the CVS at 6.5MV/cm decreases for times > 900000$s$ without showing a current increase before.

## 4. I-$Q_{inj}$ characteristics

The saturation point at $I_{max}$ is closely connected with the injected charge ($Q_{inj}$) which becomes clear from Figs. 5 and 6 when the recorded current is displayed versus $Q_{inj}$ and the saturation charge ($Q_{inj}^{sat}$) is indicated by the grey area. Figs. 5 and 6 were generated from the same data set as Figs. 1 and 4. It can be seen that $Q_{inj}^{sat}$ is constant for different fields but varies slightly for different oxide areas with $A_{ox}$=3.92$x10^{-2}$cm$^2$ in Fig. 5 and $A_{ox}$=1.64$x10^{-1}$cm$^2$ in Fig. 6. Depending on the oxide process $Q_{inj}^{sat}$ can vary significantly. Typical values from the literature are: $1mC/cm^2$ [11] and $10mC/cm^2$ [12]. For the 10.9nm oxide $Q_{inj}^{sat}$ was found to be between 10–100$mC/cm^2$.

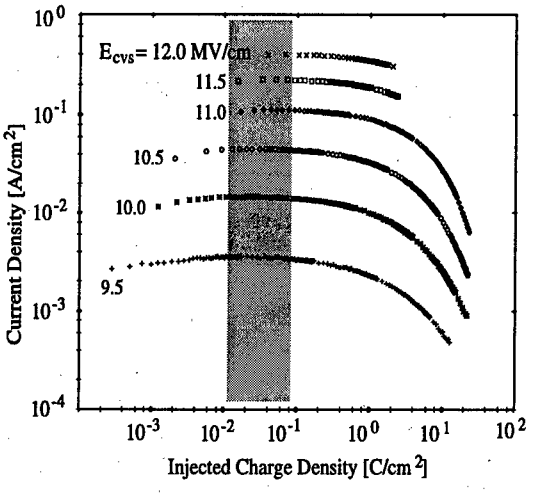



Figure 5: I-$Q_{inj}$ curves of a 10.9nm gate oxide with $A_{ox}$=3.92$x10^{-2}$cm$^2$. Breakdowns are not displayed.

Two observation can be made:

- $Q_{inj}^{sat}$ is constant for different stress fields,
- no initial current increase exists below 7.0MV/cm.

These important results support strongly Nissan-Cohen's model [2] of the finite process of initial detrapping of native electrons. They also indicate that at operating voltage the trapping characteristics of the oxide is very different to that at F-N injection.

## 5. Rate of initial charge detrapping

Another significant observation to be made from this work is that the rate of initial detrapping decreases with decreasing CVS field. The rate is defined by the change in current ($\Delta I = I_{max} - I_{init}$, which is correlated with the net trapped oxide charge [12]) versus time $t_{max}$. The rate can be

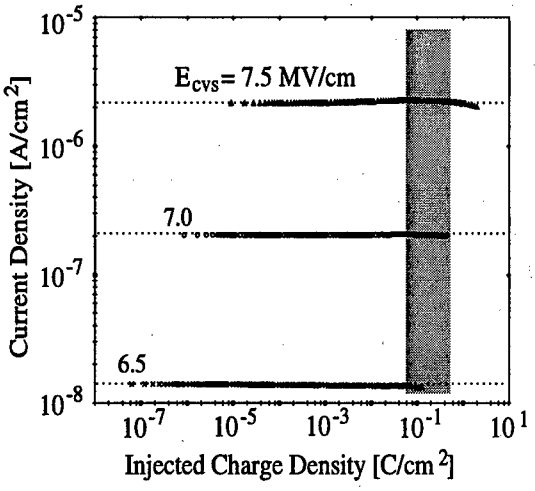



Figure 6: I- $Q_{inj}$ curves of a 10.9nm gate oxide with $A_{ox}$=1.64x10$^{-1}$cm$^2$. Breakdowns are not displayed.

experimentally determined only for I-t curves at medium and low fields which show clearly a complete initial curve trace with the initial value $I_{init}$ and the maximum current $I_{max}$. The rate is: $\Delta I/t_{max}$. Both $t_{max}$ and $\Delta$I change with stress field which is illustrated in Figs. 2 and 7. It can be seen from Fig. 7 that below 7MV/cm the current increase vanishes when the line fit of the data points crosses the zero point of $\Delta I/I_{max}$.

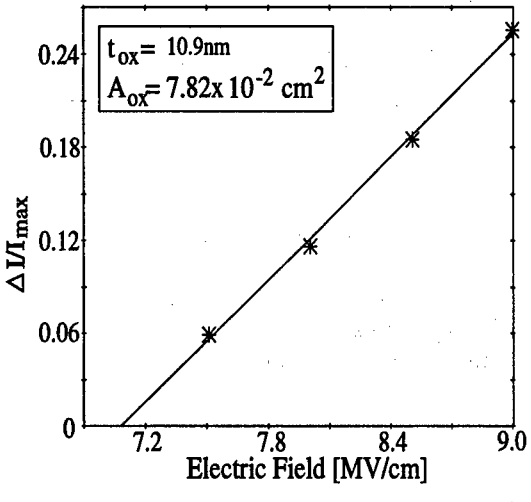



Figure 7: $\Delta$I normalised to $I_{max}$ versus $E_{cvs}$.

It has been previously shown [13,14] for CCS that a power law relationship exists between $Q_{bd}$ and $1/(\Delta V/Q_{inj})$. $\Delta V/Q_{inj}$ is the rate of charge trapping in a CCS. It has been shown [13,14] that small $\Delta V/Q_{inj}$ result in large $Q_{bd}$. Correspondingly, in this work we have found qualitatively that small $\Delta I/t_{max}$ can be related to long $t_{bd}$ which means that the rate of initial detrapping and the following trapping have a major impact on $t_{bd}$. This becomes clear when measurements are performed with CVS at two fields [10,15]. Fig. 8 shows as an example Weibull distributions of a 20nm gate oxide. The $t_{bd}$ of the two-step CVS are displayed in the right distribution with a dashed line fit where the initial detrapping/trapping is carried out at 9.0MV/cm for 5000s while the measurement is completed until breakdown at the high constant field level of 10.7MV/cm. In comparison a CVS was performed at 10.7MV/cm. Those $t_{bd}$ are shown in the left distribution with the dotted line fit. Clearly, the initial step at 9.0MV/cm results in a $t_{bd}$ (and $Q_{bd}$ which is not shown here) increase.

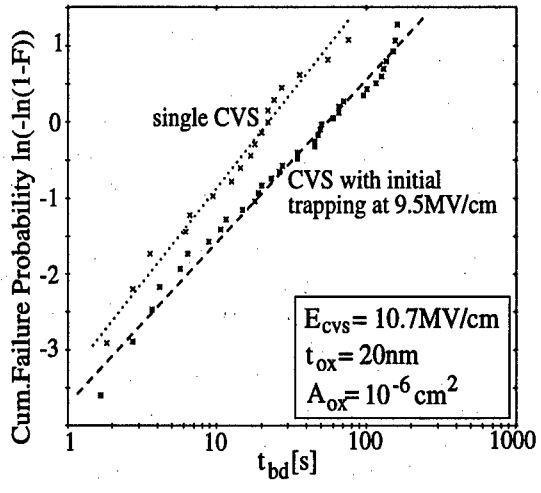



Figure 8: Two $t_{bd}$ distributions of a single CVS at 10.7MV/cm and a CVS at 10.7MV/cm with initial charge trapping at 9.0MV/cm for 5000s.

The I-t characteristics of the two-step CVS and the single CVS measurements of Fig. 8 are shown in Fig. 9. For the stress at 9.0MV/cm (lower I-t curve) it can be seen that the initial electron detrapping had been completed during the low field stress and additionally a large quantity of electrons have been trapped at the low field. This trapped charge at the low field alters the initial trapping characteristics of the CVS at high field. The I-t curve of the single CVS is steeper and starts at a higher current value.

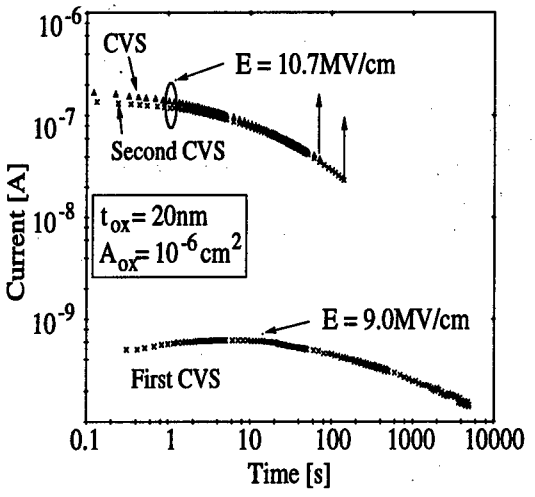



Figure 9: I-t curves of the first CVS at 9.0 MV/cm and the second CVS at 10.7MV/cm as well as the I-t curve of a single CVS at 10.7MV/cm are shown.

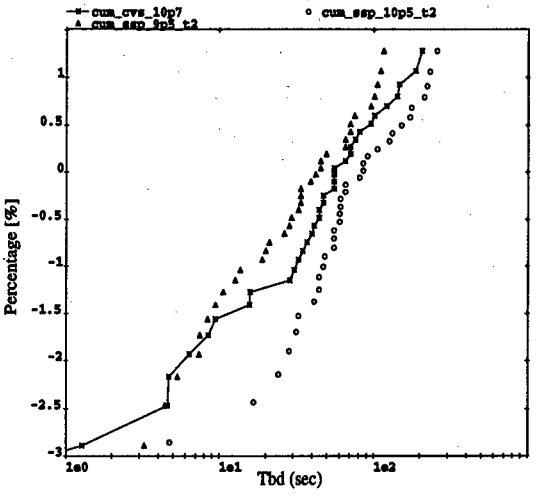



Figure 10: I-t curves of the first CVS at 9.5 MV/cm and the second CVS at 10.7MV/cm as well as the I-t curve of a single CVS at 10.7MV/cm are shown.

In contrast to this it has been found that $t_{bd}$ decreases when the low field stress ends before or near $I_{max}$. This is illustrated by the results of the following experiments. A first CVS was performed at 9.5MV/cm for 2s followed by a second CVS at 10.7MV/cm. The I-t curves of the CVS measurements at 9.5 and 10.7MV/cm are presented in in the Fig. 10. At low fields the stress was interrupted when $I_{max}$ was reached at 9.5MV/cm (see lower curve of Fig. 10). At this point the initial detrapping has not been completed and only a small quantity of electrons has been trapped so far. As a consequence the second CVS which follows the CVS at 9.5MV/cm starts at higher currents than the single CVS (two upper curves).

This has a clear impact on the $t_{bd}$ distributions of the CVS measurements at 10.7MV/cm. Fig. 11 displays the Weibull Plots of a single CVS and two CVS with initial charge trapping for 2s at lower fields. The distributions of the single CVS is displayed with a solid line. The $t_{bd}$ results of the CVS with initial trapping at 9.5MV/cm can be observed to the left which indicates a decrease of $t_{bd}$. In comparison the Weibull plot to the right presents the increased $t_{bd}$ of a CVS with initial trapping at 10.5MV/cm for 2s. The corresponding I-t curve at 10.5MV/cm (not displayed here) shows that the initial detrapping had been completed at 2s and a large amount of electrons had been trapped.

We have further obtained that for different second CVS a first fixed low CVS has different impacts. The influence of the low CVS on $t_{bd}$ increases with an increasing stress field for the second CVS. For a specific second CVS and oxide a fixed injected charge is necessary during the first CVS to observe a enhanced $t_{bd}$. This injected charge is larger than that of $Q_{inj}^{sat}$ when the electron detrapping saturates.

## 6. Conclusions and significance

It has been identified in this work that the field at which the initial de-/trapping takes place has a great impact on $t_{bd}$. Also the rate of the initial electron de-/trapping is crucial for the degradation of a gate oxide. The oxide wear out decreases for a decreasing trapping rate and for initial detrapping at low fields. Experiments show that when the initial de-/trapping had been completed at low fields the oxide exhibits an enhanced $t_{bd}$ and $Q_{bd}$. Consequently, these "wrong $t_{bd}$ have a strong influence on lifetime predictions and lead to an overestimation of oxide lifetimes at operating conditions. Below 7MV/cm an initial detrapping was no longer observed. This leads to the conclusion that at lower fields, e.g. at operating voltage, the trapping characteristics change and oxide degradation is no longer strongly influenced by the initial de-/trapping behaviour. Nevertheless, the initial

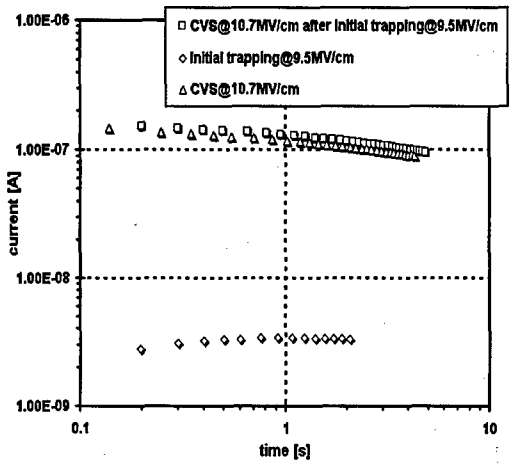



Figure 11: Two $t_{bd}$ distributions of a single CVS at 10.7MV/cm and a CVS at 10.7MV/cm with initial charge trapping at 9.5MV/cm for 2s.

de-/trapping is of great importance for highly accelerated reliability measurements.

## References

[1] Martin A. et.al., *Microelectronics Reliability*, vol.38, no.1 (1998) 37-72.

[2] Nissan-Cohen Y. et.al., *Journal of Applied Physics*, vol.60, no.6 (1986) 2024-2035.

[3] Avni E. et.al., *Journal of Applied Physics*, vol.64, no.2 (1988) 743-748.

[4] Olivo P. et.al., *Journal of Applied Physics*, vol.54, no.9 (1983) 5267-5276.

[5] Wolters D.R. et.al., *Journal of Applied Physics*, vol.58, no.2 (1985) 831-837.

[6] Yamabe K. et.al., *Journal of Applied Physics*, vol.51, no.12 (1980) 6258-6264.

[7] Liang M.-S. et.al., *Proc. of IEDM* (1986) 394-399.

[8] Duane R., The Development of Non-Volatile Memory, *M.Sc. Thesis*, NMRC, Cork, Ireland, 1996.
Duane R. et.al., *Proc. of ESREF'96* in *Microelectronics Reliability*, vol.36, no.11/12 (1996) 1623-1626.

[9] Degraeve R. et.al., *Proc. of INFOS'95* in *Microelectronic Engineering*, vol.28 (1995) 313-316.

[10] Martin A. et.al., *Proc. of IRPS* (1996) 67-76.

[11] DiMaria D.J. et.al., *Journal of Applied Physics*, vol.73, no.7 (1993) 3367-3384.

[12] Dumin N.A. et.al., *Solid-State Electronic*, vol.39, no.5 (1996) 655-660.

[13] Pesic B. et.al., *Proc. of MIEL* (1995) 273-276.

[14] Klootwijk J.H., *Proc. of ESSDERC* (1996) 369-372.

[15] Martin A. et.al., *Proc. of ESREF'96* in *Microelectronics Reliability*, vol.36, no.11/12 (1996) 1647-1650.

PERGAMON

Microelectronics Reliability 38 (1998) 1097–1101

MICROELECTRONICS RELIABILITY

# Modelling and simulation of hot-carriers degradation of high voltage floating lateral NDMOS transistors

Eric Vandenbossche, Catherine De Keukeleire, Marc de Wolf *
Hugo Van Hove and Johan Witters

*ALCATEL Microelectronics, Westerring 15, B-9700 Oudenaarde, Belgium*
**MESA Research Institute, University of Twente, Enschede, The Netherlands*

## Abstract

This paper presents the Hot Carrier Endurance of a High Voltage (100V) self aligned Floating lateral nDMOS transistor. Based on experimental results, a Safe Operating Area is determined according to maximum 10% shift of electrical parameters within 25 years. Process/Device simulation has been done in order to understand the degradation phenomena based on bulk current. Two points of high Impact Ionization rates have been found : one close to the channel junction but in depth, and the second one in the drift region. This later explains the Hot Carrier Degradation of the $R_{on}$ parameter observed experimentally.
© 1998 Elsevier Science Ltd. All rights reserved.

## I. Introduction

High Voltage (100V) MOS transistors for automotive applications are conventionally designed to operate on two bias condtions: ON state where gate bias is high but drain bias is low, and the OFF state where gate bias is 0V and drain bias is as high as possible. The presence of a high series resistance, namely the drift region, allows to handle such high drain bias. However, during switch between ON and OFF states, both high gate and high drain bias occur. This results in electrical parameter degradation due to Hot Carrier Injection, HCI. This paper presents the analysis of the HCI behavior of a high voltage floating nDMOS transistor, FnDMOS. In addition, Process/Device simulation has been performed to understand the physical reasons of electrical parameter degradation. The simulation has been fully calibrated to experimental results, both main electrical parameters and bulk current under various bias conditions. Therefore, a Safe Operating Area, SOA, is determined based on analyzing Hot-Carrier bulk current and the relative degradation. A comparable behavior of electrical parameter degradation has been found on a submicrometer LDMOS transistor [1].

## II. Device description

The High Voltage FnDMOS is processed in a 0.7μm CMOS technology. Figure 1 shows a typical cross section of the device. The Channel is performed by a lateral diffusion of a PBODY boron implant inside a NTUB layer that makes the drift region. The NWELL added on the drain contact allows to support high voltages above 100V. The gate oxide thickness is 42nm. All processing steps related to High Voltage devices are realised prior to processing CMOS transistors in order to keep the

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00115-2

electrical compatibility with the standard digital 0.7μm CMOS technology. Typical electrical parameters of the FnDMOS are presented in Table I.

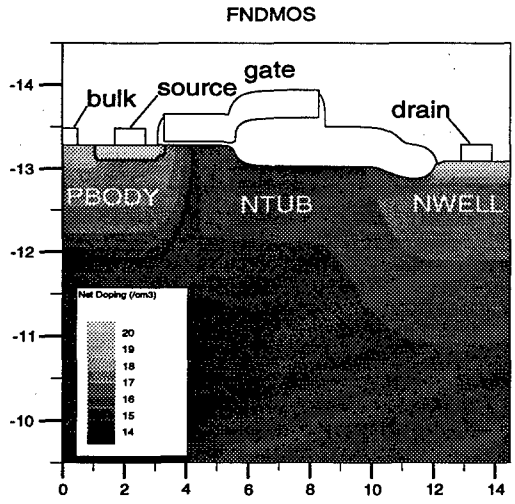



Figure 1 Cross section of the High Voltage Floating nDMOS transistor from process simulation. Iso contours of net doping concentration in silicon are shown.

Table I Electrical parameters of the High Voltage Floating lateral nDMOS transistor. The values in brackets are standard deviation. Simulation results are shown.

| | $V_{th}$ (V) | β (μA/μm/V) | $I_{dlin}$ $V_d$=0.1V $V_g$=12V (μA/μm) |
|---|---|---|---|
| Exper. | 2.46 (0.08) | 21.72 (1.0) | 3.2 |
| Simu. | 2.42 | 20.94 | 3.3 |

| | $R_{on}$ (kΩ.μm) | $I_{dsat}$ $V_d$=40V $V_g$=3.5V (μA/μm) | $I_{dsat}$ $V_d$=40V $V_g$=5V (μA/μm) | $V_{BD}$ (V) |
|---|---|---|---|---|
| Exper. | 33.56 (1.26) | 12.5 (0.64) | 54 (1.41) | 104.6 (2.77) |
| Simu. | 31 | 13.3 | 54.9 | 104.5 |

## III. Results

### *Experimental set-up*

An experimental set-up was developed to stress high voltage DMOS transistors up to 200V. A parameter analyzer HP4156A with an extension module (allowing stress up to 200 V) is used to stress and measure the devices. 8 devices can be stressed at the same time ; the switching for measurements is done with a switching controller HP4083A. A computer controls the test bench via HPIB interface.

A measurement program written in HPVee controls the measurements of transistor parameters ($V_t$, β, $I_{dsat}$ and $R_{on}$) during stressing.

With high voltage transistors, one issue is the large temperature increase during stress. This induces an error on the measured parameter because the temperature of the transistor is unknown and fluctuates with time. Therefore it was important to find an approach where the temperature effect could be compensated. Monitoring the subthreshold slope solved this issue : when slope variations are below 0.2%, then the transistor parameters can be measured.

### *Hot carrier degradation model*

Purpose of hot carrier degradation measurements is to determine the Safe Operating Area where the lifetime for 10% degradation of any parameter of the transistor is above 25 years.

$R_{on}$ was found to be the faster degraded parameter. Various stress conditions were chosen :

- from 80 to 100V on the drain at a constant gate voltage of 12V
- from 4 to 12V on the gate at a constant drain voltage of 100V.

As expected at constant gate voltage, the higher the drain voltage the faster the degradation. But hot carrier degradation at constant drain voltage is more complex as illustrated by Figure 2 where the time of 10% $R_{on}$ shift (in log scale) is drawn as a function of $V_g$. Two behaviours are detected : first, typical increase of degradation with increase of gate voltage (from 0 to 8 V), and second, decrease of degradation when stress voltage increases (from 8 to 12V).

These two effects will be explained below with simulation.

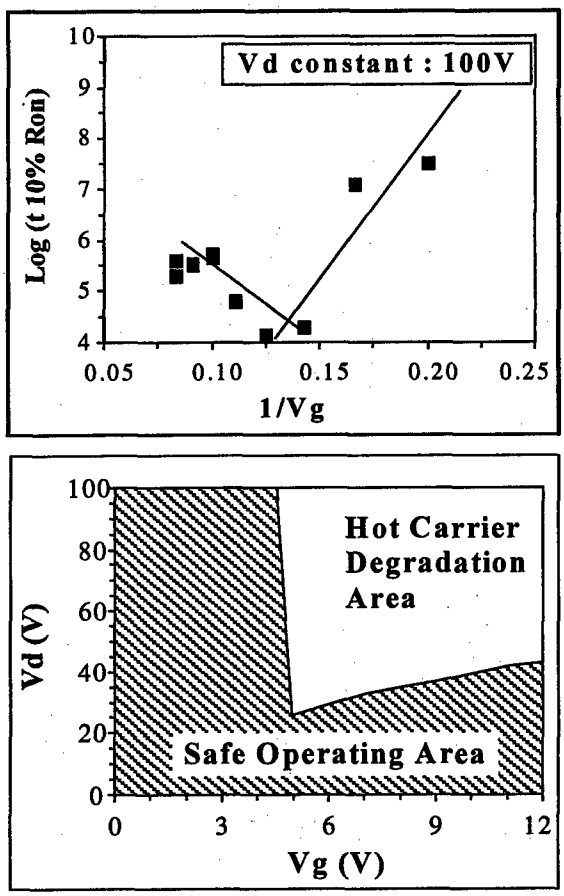



Figure 2 Time of 10% Ron shift as a function of Vgs (top). SOA for a maximum of 10% Ron degradation over 25 years (bottom).

A lifetime model including both $V_d$ and $V_g$ is used to fit the experimental data. Two equations based on Takeda model were extracted depending on the gate voltage range:

- for $V_{gs}$ between 0 and 8V :

$$Log_{10}(\tau) = \frac{50.9}{V_g} - 2.3 \quad (1)$$

- for $V_{gs}$ between 8 and 12V :

$$Log_{10}(\tau) = \frac{237.3}{V_d} - \frac{29.3}{V_g} + 5.9 \quad (2)$$

The experimental SOA for a maximum of 10% $R_{on}$ degradation over 25 years is then derived from equations (1) and (2) as illustrated in Figure 2.

## IV. Simulation

### *Process/Device simulation calibration.*

The process simulation tool used for the study is IMPACT4 (Isen Modelling Package for Circuit Technology) [2]. The device simulation tool is ATLAS [3]. Table 1 presents simulation results in comparison with experiments. A very good agreement has been obtained. Both bird's beak topology of isolation structure (LOCOS) and dopant diffusion and OED of PBODY, NTUB and NWELL layers have been calibrated. For device simulation, the CVT mobility has been chosen. Degradation parameters of mobility due to vertical electrical field have been calibrated. In addition, the breakdown voltage simulation has been done with using the Impact Ionization model (Selberherr model) with a modified critical electrical field for holes. Based on this calibration, it allows to understand the Hot Carrier Degradation of the FnDMOS observed experimentally.

### *Bulk current simulation.*

Figure 3 presents bulk current as a function of gate bias and ratio between bulk to drain currents as a function of drain bias. Simulation results are also shown where a good agreement is obtained with experiments. Figure 4 shows a cross section of the device at $V_{DS}$=60V and 3 values of gate bias. The gate bias are chosen so that the first corresponds to maximum bulk current at low gate bias, the second one is where bulk current has a minimum above its maximum, and the third at maximum gate bias. The Impact Ionization (II) rate is shown for the 3 bias conditions. For low gate bias and at high $V_{DS}$ (above $V_{sat}$), the current flows into the NTUB since the drift region close to the surface is pinched off due to high vertical electrical field (saturation regime). In this regime, the FnDMOS acts as a standard MOS where the maximum bulk current is generated in the area of the PBODY/NTUB junction (location where both high electrical field and high current density occur). By increasing the gate bias, the local $V_{GD}$ in between the edge of the PBODY/NTUB junction and the bird's beak reduces lowering the local vertical electrical field. Then this region is less and less pinched-off such that electrons flowing from the

source to drain come back to the surface at high gate bias.

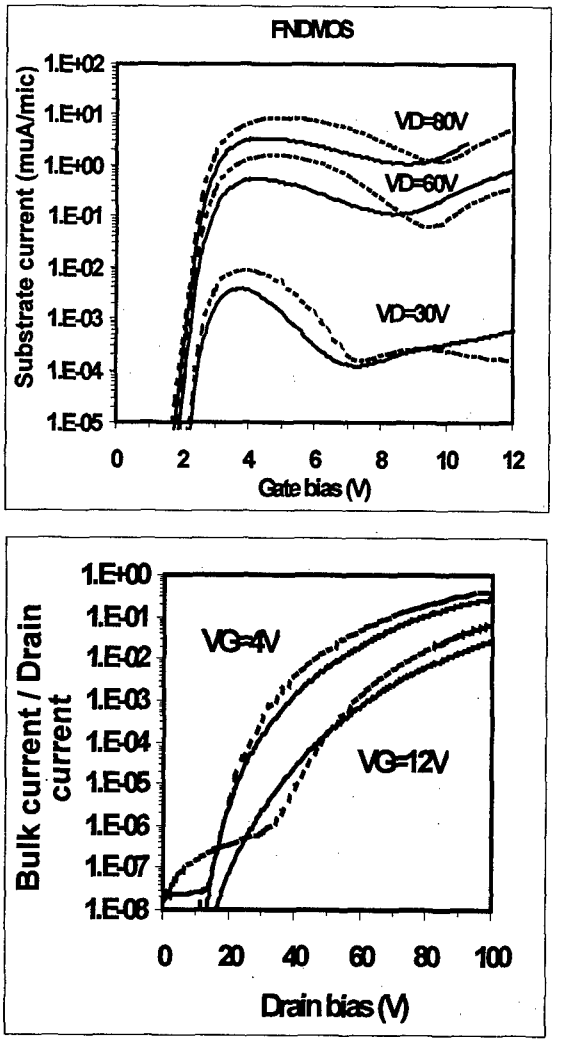



Figure 3 Substrate current as a function of gate bias (top) and ratio of bulk to drain currents as a function of drain bias (bottom) from experiments (full lines) and simulation (dashed lines).

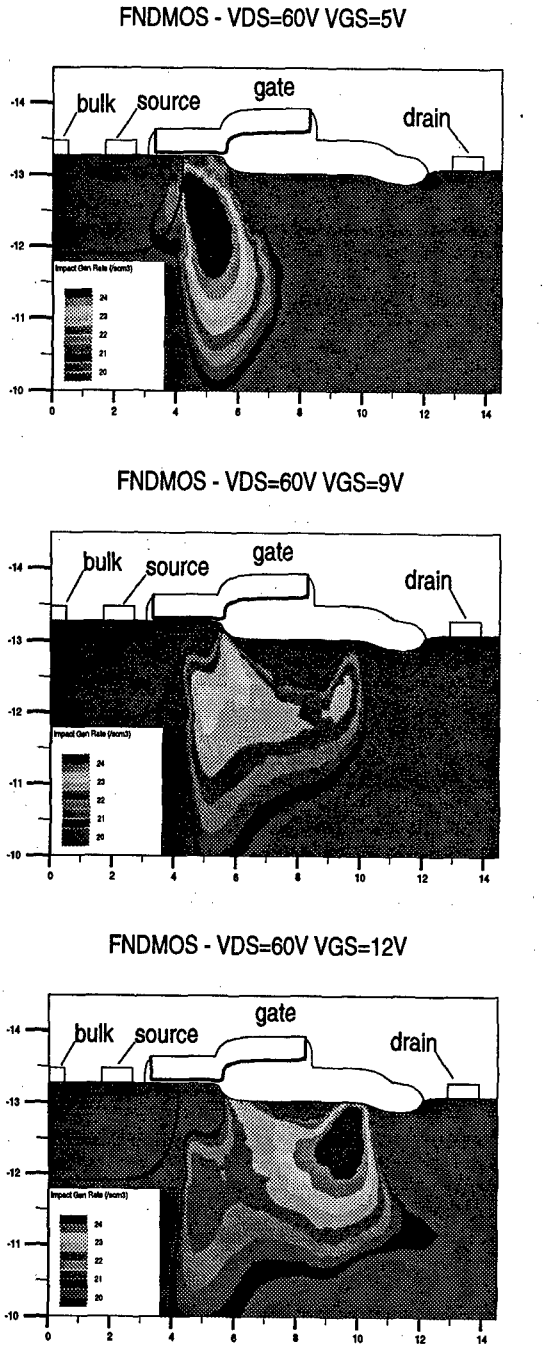
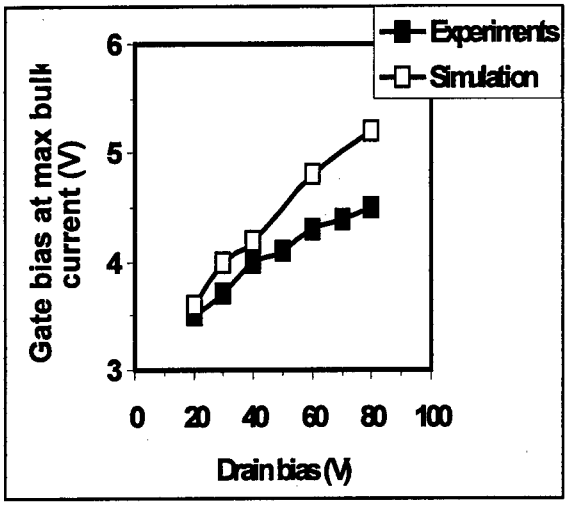



Figure 4 Cross section of the lateral FnDMOS. Iso contours of Impact Ionization rate are shown for a fixed drain bias (60V) and various gate bias.

According to this phenomenon, the maximum electrical field is displaced from the PBODY/NTUB junction into the NTUB drift region. Thus, a second point of Hot Carrier generation occurs in the NTUB/NWELL region. The transition between these two locations is mostly function of gate bias, while drain bias affects only the II rate magnitude (Figure 4). A comparable behavior has also been reported for submicron LDMOS transistor [1].



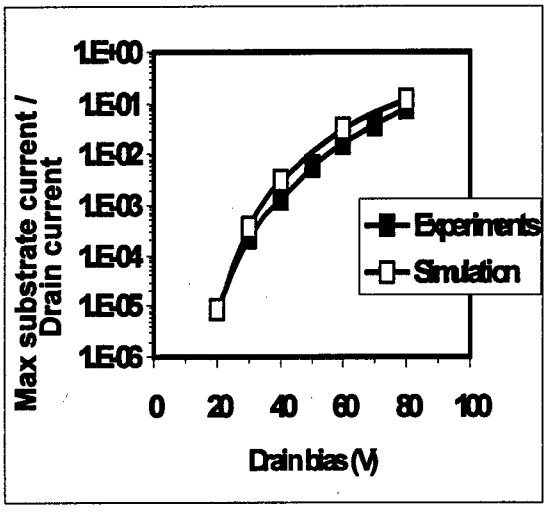




Figure 5 Gate bias at maximum bulk current (top) and ratio of maximum bulk current to drain current (bottom) as a function of drain bias.

Figure 5 presents the gate bias and the ratio of bulk current to drain current at maximum bulk current as a function of drain bias. A good agreement is found between simulation and experiments. The gate bias at maximum bulk current does not follow the rule observed on standard MOS transistor (around half of drain bias). This is linked to the drift region (high access resistance) and to the complex behavior of this region at various bias conditions from either gate or drain.

Since the first location of the maximum II rate is far away from the silicon/gate oxide interface, there is very little degradation of $V_t$, $\beta$, $I_{dsat}$ and $R_{on}$ at low gate bias as observed experimentally. While the second location of maximum II rate, close to the Si/SiO2 interface, results in high degradation of only $R_{on}$. Therefore, the SOA obtained from the experiments is explained by the second location of the maximum II rate where Hot Carrier Injection takes place.

## Conclusion

Hot Carrier degradation behavior of electrical parameters has been analyzed for a High Voltage FnDMOS transistor. Based on experimental results, a Safe Operating Area, SOA, has been deduced. Process/Device simulation has been extensively used in order to understand the experimental observations. Simulation results have been calibrated to experimental data and also bulk current in various gate, drain bias conditions. Two different locations where Hot Carriers are generated were found. The transition between these two locations is only gate bias dependent. Also, the worst degradation has been observed at high gate bias that is explained by a high Impact Ionization rate in the drift region (second location). The $R_{on}$ parameter has been found to be the faster degradation parameter. Since bias conditions for measuring this parameter are the same as the one for ON state, the physical understanding of the degradation mechanisms is of very importance to allow the optimization of this device in order to increase the SOA.

## References

[1] R. Versari *et al.*, "Hot Carrier Reliability in Submicron LDMOS transistors", IEDM'97 Washington D.C. and related references.
[2] IMPACT4.8 Manual 1997.
[3] SILVACO package manual 1997.

PERGAMON

Microelectronics Reliability 38 (1998) 1103–1107

MICROELECTRONICS RELIABILITY

# A new hot carrier degradation law for MOSFET lifetime prediction

B. Marchand[a], G. Ghibaudo[a], F. Balestra[a], G. Guégan[b] and S. Deleonibus[b]

[a] *Laboratoire de Physique des Composants à Semiconducteurs (UMR CNRS/INPG) ENSERG-INPG, BP 257, 38016 Grenoble Cedex 1, France*
[b] *LETI-CEA, 17 rue des Martyrs, 38054 Grenoble Cedex 9, France*

## Abstract

We propose in this paper a new hot carrier degradation law for a reliable MOSFET lifetime prediction. We show that the proposed exponential function can describe all kind of curve concavity (saturating or non-saturating shapes) and can fit very well with the experimental data for the whole duration of the stress. Finally, it gives a more accurate lifetime value as compared to previous modelings because it accounts for the concavity of the saturating degradation law.
© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Device reliability is a great concern in microelectronics research. Hot carrier stresses are carried out in order to forecast the evolution of degradations in the device, monitoring a large amount of electrical parameters which interest circuit designers. Lifetime prediction defined as the time at which a parameter reaches its maximum acceptable shift makes up the main result in terms of reliability for a technology.

Several extrapolation laws are used to calculate the lifetime. The first one proposed in the 80's is the power law [1]. In fact, with MOSFET downscaling, a break in the law appears during the stress. In log-log scale the straight line related to the power law often disappears and gives way to a curved degradation law with saturation. Some authors have tried to split up the kinetics into two successive power laws [2]. But, the splitting is critical and in a way arbitrary. Other authors have proposed a logarithmic law which should fit better [3].

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00137-1

## 2. A federative law

In this paper, we present a federative law which consists of the following exponential relationship : $J=J_0.\exp(-B.t^{-n})$. It has been first found in stress-induced leakage current (SILC) analysis [4]. We show here that it is also fully valid for hot carrier degradation. The exponential function has got a major advantage : it describes all type of curve concavity depending on the sign of the exponent n. We speak about a saturating exponential law when n is positive and a non-saturating one in the other case. Our function is no more valid in the case of n equal to zero because it becomes time independent and only a power law can describe the degradation. However, the exponential function fits quite well with the parameter shift for an exponent n close (but not equal) to zero. Factor B adjusts the kinetics dynamics and factor $J_0$ the amplitude.

In figure 1, four theoretical degradation laws are plotted formatted like experimental data. We consider the power, logarithmic, saturating exponential and non-saturating exponential laws.

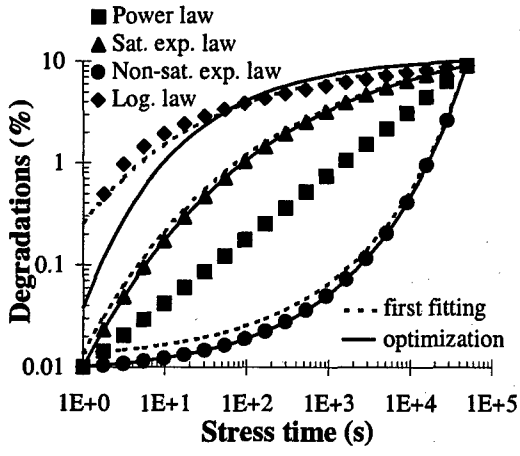



Fig. 1 : Arbitrary degradation laws and their fitting with the exponential law.

These functions describe a degradation equal to 0.01 % at 1 s stress time and to 10 % at $5.10^4$ s stress time. We calculate the logarithmic derivative of each one as done in a SILC analysis and multiply them by the time. These calculated functions, that we can called degradation rates (no physical dimension), are plotted versus stress time in the discrimination plot (see Fig. 2).

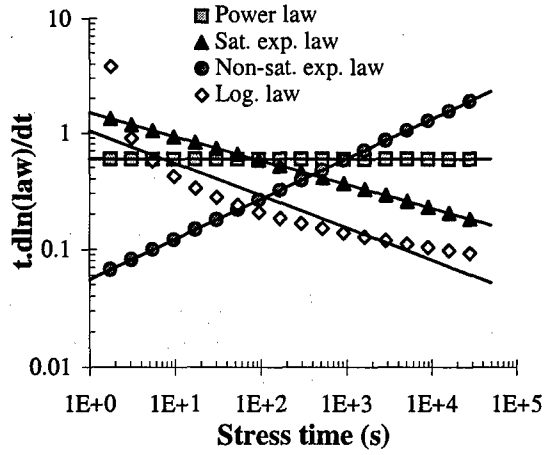



Fig. 2 : Degradation rates to extract the parameters of the exponential law.

Here is the relationship we get for the exponential law : $t.d\ln(J)/dt=n.B.t^{-n}$. This function is a straight line in a log-log scale which slope gives the exponent n and quickly informs about the saturating (negative slope) or non-saturating (positive slope) nature of the law. A flat curve means a power law. The shape of the logarithmic law is distinguishable from the others by an upward bend for short stress times. For each function we extract the parameters n (opposite of the slope) and B (degradation level at 1 s stress time divided by n) of the exponential law and calculate $J_0$ in order to make coincide with the final data point as we want afterwards to extrapolate towards longer stress times. After that, their fittings are drawn in the degradation space as dash lines (Fig. 1).

We can verify that, particularly for long stress times, there is a good agreement between the theoretical degradation laws and their fittings with the exponential law. However, we can optimize the fitting using for instance a least-squares algorithm (solid lines in figure 1).

We can notice that the optimization fails for the logarithmic law as there is an abrupt variation of the law for the first two points. In this case, one should prefer the first parameter set which makes coincide with the final data point.

Finally, we call attention to the following critical point : the stress time step has to be accurately logarithmic and constant in order to prevent from calculation noise.

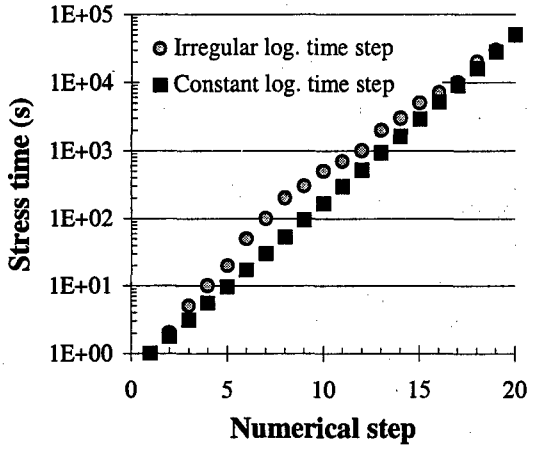



Fig. 3 : Relationship between stress time and numerical step. Impact of the logarithmic time step.

Indeed, an irregular step (for instance, measurements at 100, 200, 300, 500, 700 s instead of 100, 158, 251, 398, 631 s) induces a non-linearity in logarithmic scale between the stress time and the numerical step (see Fig. 3).

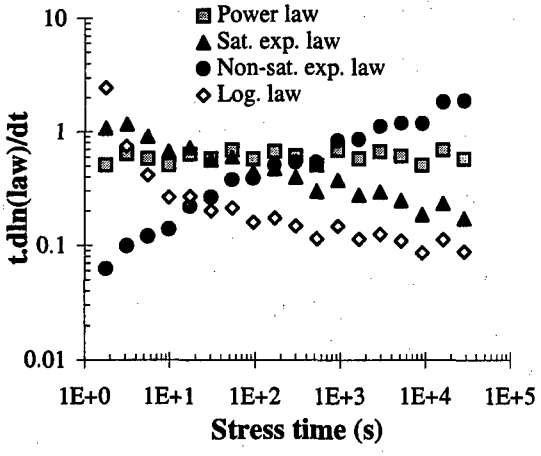



Fig. 4 : Degradation rates. Calculation noise due to an irregular logarithmic time step.

This non-linearity has repercussions on the derivative of the laws with respect to the time and gives noisy results (see Fig. 4). This noise would be added to the measurement noise, which has to be avoided.

## 3. Application to measurements

In figures 5 to 10, we consider experimental data from two hot carrier stresses. One was carried out on a 0.5 μm n-MOSFET at the maximum gate current (Vd=2.5V, Vg=1.6V, Vb=-5V) ; the other one on a 75 nm n-MOSFET from the same technology at the maximum substrate current (Vd=2.5V, Vg=0.8V, Vb=0V). In order to present various electrical parameters, we have arbitrarily chosen to draw the degradation of the saturation drain current (measured in direct mode Isatd and in reverse mode Isat) for the 0.5 μm n-MOSFET, and the degradation of the threshold voltage Vtle (obtained by linear extrapolation) and of the maximum transconductance Gmmax for the 75 nm n-MOSFET.

In figure 5, the straight lines for the shifts of the parameters show that the degradation follows a saturating exponential law as previously defined.

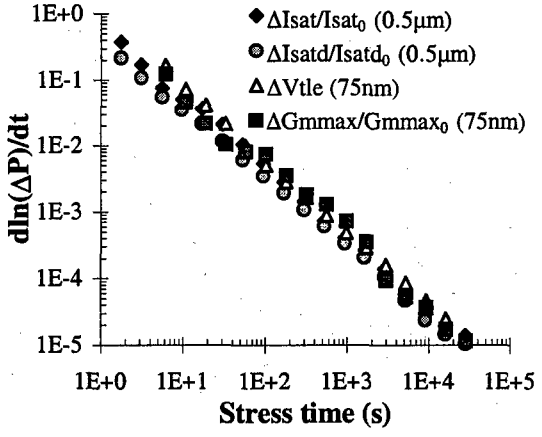



Fig. 5 : Logarithmic derivatives of the degradations for various parameters.

The discrimination plot (see Fig. 6) informs about the trends of the degradation laws. While the threshold voltage degradation rate almost describes a straight line, the transconductance one only shows a saturating exponential trend from which we can extract parameters n and B.

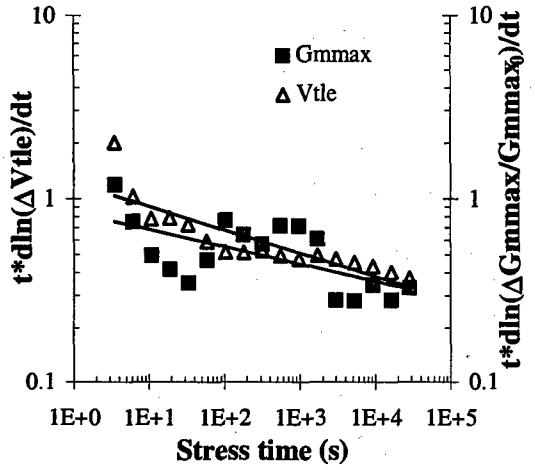



Fig. 6 : Discrimination plot for the parameter extraction. 75 nm n-MOSFET.

Despite the transconductance noise, it fits quite well in the degradation space (dash lines in Fig. 7) as it does for the threshold voltage (dash lines in Fig. 8). Then, we adjust the parameter set with a computerized fit optimizer (solid lines in Figs. 7 and 8).

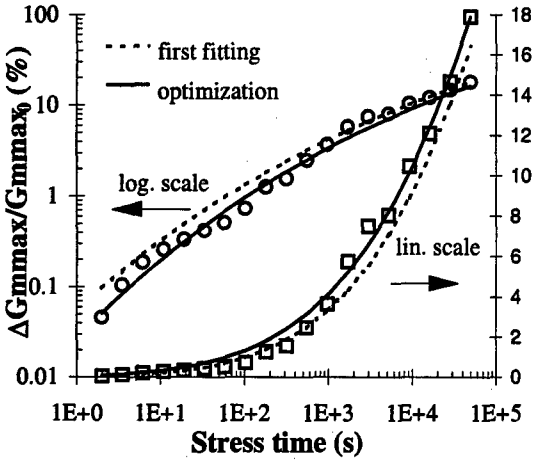



Fig. 7 : Fitting for the maximum transconductance. 75 nm n-NMOSFET.

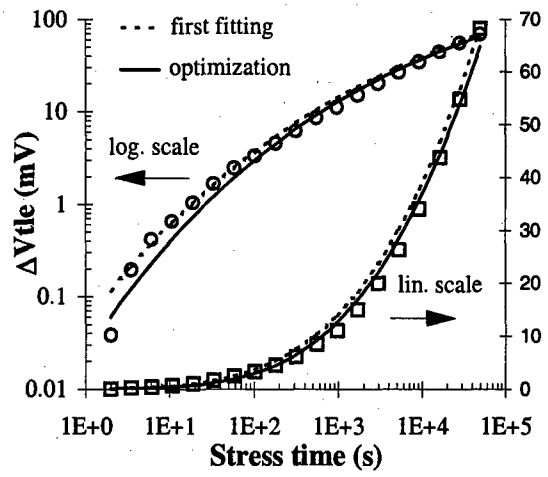



Fig. 8 : Fitting for the threshold voltage. 75 nm n-MOSFET.

In figure 9, the extraction of the exponential law parameters also results in a good agreement with the experimental data. The determined exponential law enables us to calculate the lifetime according to a 10 % degradation criterion and we compare with the result obtained by the use of a power law extrapolation.

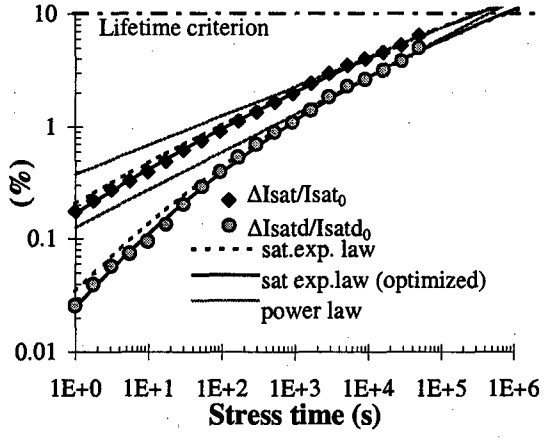



Fig. 9 : Comparison between the parameter degradations and their fittings with the saturating exponential law (optimized and not) and the power law. 0.5 μm n-MOSFET.

Figure 10 points out that the extrapolation based on the saturating exponential law yields a more optimistic evaluation of the lifetime than

with the power law. Indeed, it takes into account the concavity of the degradation law.

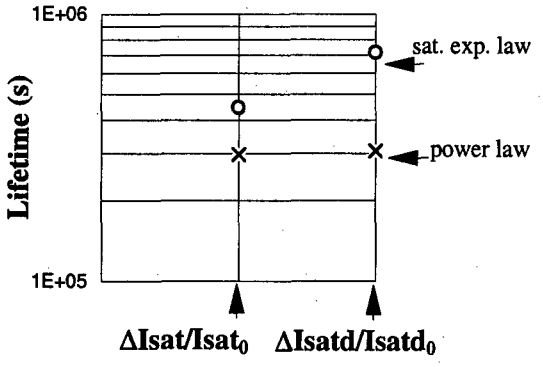



Fig. 10 : Comparison between power law and saturating exponential law extrapolation. 0.5 µm n-MOSFET.

## 4. Conclusion

To sum up, we have proposed in this paper to apply and generalize the exponential function used in SILC analyses to hot carrier studies. We have shown that this exponential function can describe all kind of curve concavity (saturating or non-saturating shapes) and can fit very well with the experimental data for the whole duration of the stress. Finally, it gives a more accurate lifetime value because it accounts for the concavity of the degradation law.

## References

[1] E. Takeda and N. Suzuki, IEEE Electron Device Letters, vol. EDL-4, p. 111, 1983
[2] V.H. Chan and J.E. Chung, IEEE Trans. on Electron Devices, vol. 42, pp. 957-962, 1995
[3] S.H. Renn, J.L. Pelloie and F. Balestra, Proc. ESSDERC'97, p.696, 1997
[4] A. Scarpa et al., Electronics Letters, vol. 33, p. 1342, 1997

PERGAMON

Microelectronics Reliability 38 (1998) 1109–1113

MICROELECTRONICS RELIABILITY

# Recovery and stress dynamics in bipolar transistors and MOS devices

F. Ingvarson, L-Å. Ragnarsson, P. Lundgren

*Department of Microelectronics, Solid State Electronics Laboratory, Chalmers University of Technology, S-412 96 Göteborg, Sweden*

**Abstract**

Stress and recovery dynamics of bipolar transistors and ultra thin oxide MOS devices have been investigated. We have found that these devices can exhibit similarities in the stress dynamics. The recovery during heat treatment was also investigated and it was found that both the dynamics and the temperature dependence of the recovery were very similar for both bipolar and MOS devices. These findings indicate that the defects might be similar where bipolar current gain degradation and MOS gate oxide charging are concerned.

© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

In BiCMOS circuits where bipolar transistors are interfaced with CMOS transistors, the bipolar transistors can experience a reverse base-emitter bias when the CMOS transistors are switched [1, 2]. This causes a degradation of the current gain with time due to an increase in the base current, which affects both speed and long-time reliability of the circuit. This degradation also puts restrictions on both transistor design (e.g. high breakdown voltage to increase device reliability) and circuit design (e.g. low operation voltage, $V_{BE}$>-2.5 V).

The cause for the increase in base current is believed to be a degradation of the $Si/SiO_2$ interface at the base-emitter p-n junction [3]. One mechanism that could account for this degradation is hot carrier injection into the oxide; high electric fields due to the highly doped emitter and base regions can generate hot carriers which in turn can generate interface traps and/or cause carrier trapping in the oxide.

The current gain degradation has traditionally been monitored by investigating the increase in forward operation base current where the stressing is conducted by forcing a constant reverse base current [4, 5] or applying a reverse voltage [6, 7] with the collector left open. The collector current is unaffected by the stressing. From these measurements the transistor's time-to-failure can be estimated [8].

The recovery of the base current has been studied using high temperature annealing [5, 6] and forward biasing of the base-emitter junction [9, 10, 11].

In our study we have investigated stress and recovery dynamics in both bipolar transistors and thin oxide MOS devices to gain further information about the nature of the degradation of these devices. In the stress experiments we have monitored the *reverse* base current in the bipolar transistor and the direct tunnel current through the gate oxide in the MOS device. The recovery is investigated using high temperature annealing. In contrast to Huang [6], we investigate the recovery dynamics for the initial relaxation, spanning four orders of magnitude in time, ranging from 1-10000 s.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00088-2

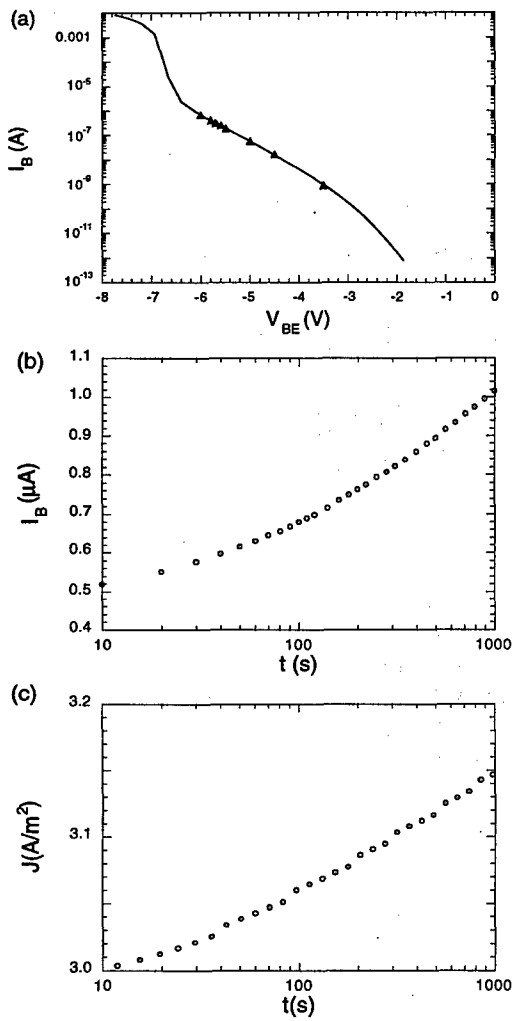



Fig. 1. (a) Reverse base current vs. $V_{BE}$ with the stress voltages indicated. (b) Bipolar stress transient with $V_{BE\text{-stress}}$= -5.5 V. (c) NMOS stress transient (oxide thickness ≈ 2.5 nm) with $V_{G\text{-stress}}$=1.8 V.

## 2. Experimental procedures

Commercially available npn-type bipolar transistors with an emitter area of 1x30 μm$^2$ from a 0.8 μm BiCMOS process were stressed at room temperature (RT) by applying a constant base-emitter voltage $V_{BE\text{-stress}}$ in the range -3.5 V to -6.0 V and forcing a zero collector current for 1000 s. The stress voltages were chosen to ensure that the reverse base current was in the tunnelling regime. Fig. 1a shows the reverse $I_B$-$V_{BE}$ characteristic with the stress voltages indicated. In Fig. 1b the stress transient for $V_{BE\text{-stress}}$=-5.5 V is shown.

After stressing, a forward RT Gummel characteristic ($I_B$-$V_{BE}$) was acquired under a constant collector-emitter voltage to define the initial post-stress condition of the device. Transistors stressed at $V_{BE\text{-stress}}$=-5.5 V were then annealed at the temperatures $T$=RT, 100, 170, 240, and 310 °C, respectively, for the cumulative times t=1+9+90+900+9000 s. Each transistor was evaluated at each time step by obtaining the RT forward Gummel characteristics. The base-emitter voltage was kept low enough to prevent recovery from forward biasing.

The MOS devices used in the stress experiments were fabricated from <100> oriented, 1-50 Ωcm substrates of n- and p-type. For the p-type devices the gate material was n-type poly silicon whereas for the n-type devices aluminium evaporated from a resistively heated crucible was used. The oxide thickness was 2.8 nm (p-type) and 2.5 nm (n-type) after post metallization anneal (*PMA*), as extracted from capacitance-voltage (*C-V*) data using a method described by Maserjian [12]. The final processing step was a *PMA* to improve oxide integrity [13].

A gate to substrate voltage in the range 1.0 V to 2.5 V for n-type devices and -4.7 V to -5.0 V for p-type devices was used to stress the samples for 1000 s at RT. Fig. 1c shows the stress transient for an n-type device with $V_{G\text{-stress}}$=1.8 V.

The MOS devices used in the recovery experiments were fabricated on p-type substrates with an oxide thickness of 2.3 nm. The gate material was aluminium. Stressing was carried out at $V_{G\text{-stress}}$=-3 V for 50 s. This low stress time was chosen to prevent fatal device breakdown during stress. Immediately after the stress the device was allowed to recover at -1.2 V for 100 s. This scheme reduces the influence of the substantial recovery that takes place during current-voltage measurements. After this stressing/recovery treatment a current-voltage (*I-V*) curve was measured, and this curve defined the initial post-stress condition of the device. Further *I-V* curves were then measured after the pre-set times of annealing.

## 3. Results and discussion

### 3.1. Stress dynamics

It has been suggested that base current degradation as a result of interface state generation and oxide charging is due to hot holes originating from the tunnelling of base valence band electrons to empty

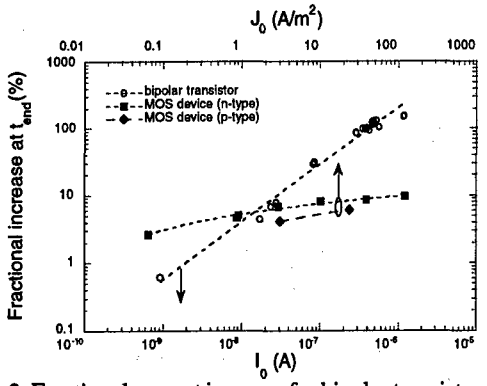



Fig. 2. Fractional current increase for bipolar transistors and MOS devices at $t_{end}$≈1000 s.

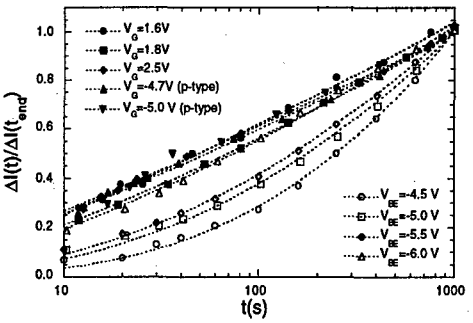



Fig. 3. Stressing dynamics of bipolar transistor (open markers) and MOS device (filled markers) for different stress voltages.

emitter conduction band states [2, 14]. Assume that all interface traps generated or charge trapped in the oxide gives an equal amount of increase in the reverse base tunnelling current $I_{Br}$. The fractional current increase is then proportional to the density of defects $N_T$ (interface traps or oxide charge)

$$\frac{\Delta I_{Br}(t)}{I_{Br0}} \propto N_T(t) \tag{1}$$

where $I_{Br0}=I_{Br}(0)$. For MOS devices we have previously shown that electrical stress of the MOS devices lead to the appearance of positive oxide charge inducing an increase in the direct tunnel current which is proportional to the oxide charge density [15]

$$\frac{\Delta I_G(t)}{I_{G0}} \propto N_{OT}(t) \tag{2}$$

where $I_{G0}=I_G(t=0)$ and $N_{OT}$ is the density of positive ions trapped in the oxide.

In Fig. 2 the fractional increase of the current at $t_{end}$≈1000 s is depicted for different initial currents. For the bipolar transistors, the relationship between initial current and subsequent current increase during stress holds both when changing the initial current by applying different stress voltages and when looking at devices with different initial current at the same stress voltage. However, as was shown in a previous study [15], the initial current alone is not able to predict the charging of MOS devices with different oxide thicknesses. The voltage across the oxide or rather the oxide field seems to depict the increase in current more accurately. This is evident from Fig. 2 where the p-type MOS devices show a slightly smaller fractional current increase at equal initial current density as compared to the n-type devices. This is a result of the different gate material and oxide thickness. We have not been able to verify that the same kind of dependence on the electric field exists in the bipolar case. In our experiments we have only access to one type of bipolar transistors regarding the emitter and base doping concentrations. Hence we cannot choose the initial current and the applied stress voltage independently of each other.

The stressing conditions in the two types of devices are quite different due to the different device geometries. In the bipolar transistor, the potentially damaging motion of hot carriers is along the $Si/SiO_2$ interface, while in the MOS device it is transversal to the gate oxide. Also, due to the differences in applied voltage, the carriers in the bipolar transistors can become more energetic than the tunnelling electrons in the MOS device. The equivalent area of bipolar transistors is unknown and we therefore choose to show the increase with current for the transistors, whereas for the MOS devices we use the current density.

It is clear that for both types of devices there is a correlation between the initial current and the fractional increase of the current. For the n-type MOS devices the fractional increase ranges from approximately 3% to 10% when the initial current increases with 3 orders of magnitude whereas for the bipolar transistors it increases from 0.6% to 150%. This large difference needs not be evidence of a difference in the degradation mechanism between bipolar transistors and MOS devices since the impact of the charges/defects rather than their density can have a strong correlation with the current in the case of the bipolar transistors.

The stressing dynamics can be investigated by plotting

$$\frac{\Delta I(t)}{I_0} \Big/ \frac{\Delta I(t_{end})}{I_0} = \frac{\Delta I(t)}{\Delta I(t_{end})} \quad (3)$$

which is the fractional increase in the current normalized to the total fractional current increase at $t=t_{end}$ where $t_{end}$ is the time when the stress was interrupted. Fig. 3 shows the stressing dynamics for both bipolar transistors and MOS devices. For bipolar transistors there is a clear voltage dependency on the stress dynamics, whereas in comparison, the MOS device stress dynamics are relatively unaffected by the stressing voltage. Notice that for high $V_{BE\text{-stress}}$, the transistor and MOS device stress dynamics coincide.

### 3.2. Recovery dynamics

Fig. 4 shows the current recovery expressed as $\Delta I(t)/\Delta I_0$ where $\Delta I(t)=I(t)-I_{\text{unstressed}}$ is the current increase at time $t$ and $\Delta I_0=I_{\text{initial post-stress condition}}-I_{\text{unstressed}}$ is the initial current increase. In the bipolar case $I$ is the forward base current measured at a base-emitter voltage of 0.55 V and a collector-emitter voltage of 3 V. In the MOS case $I$ is the direct tunnel current through the gate oxide measured at a gate substrate voltage of -1.2 V.

Using the same assumption used in Eq. 1 it follows for the forward base current that

$$\Delta I_{Bf}(t) \propto N_T(t). \quad (4)$$

The RT exposure of the transistors during measurements does not have to be considered since the RT recovery is negligible for $t<1000$ s. However, the RT recovery of the gate current is substantial and gives a reduction of the oxide charge density of about 5% after 100 s. Since the annealed samples are exposed to approximately 100 s at RT, the measured amount of remaining defects is adjusted up by 5 percentile units in order to take the RT recovery into account. Due to the uncertainty of this correction, and due to the fact that there is a relatively large spread among devices, in particular for short annealing times (where the accuracy of timing is estimated at ±50%), the uncertainty in the value is rather large, estimated at ±0.05 in $\Delta I(t)/\Delta I_0$. The result of the 310°C heat treatment after stress is strongly influenced by effects occurring also in unstressed devices at this temperature. Therefore we are unable to draw any conclusions regarding the effect of heat treatment on the stress induced damage

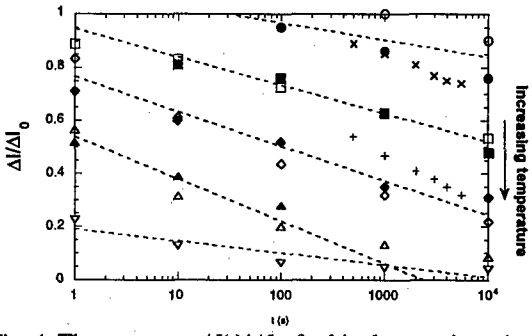



Fig. 4. The recovery, $\Delta I(t)/\Delta I_0$, for bipolar transistor (open markers) and p-type MOS device (filled markers) for the temperatures $T$=RT, 100, 170, 240, and 310°C (bipolar only). Dashed lines are only a guide for the eye. Data showing the shift in flat band voltage, $\Delta V_{FB}(t)/\Delta V_{FB}(0)$, from Lakshmanna and Vengurlekar [16] is also shown (x=28°C, +=103°C).

at this temperature.

Note the similarity in the recovery dynamics and temperature dependence of the bipolar transistor and the MOS device. Hence it is likely that the passivation mechanisms are the same in the two cases further indicating that the type of defects are the same. This is further supported by the fact that the two types of devices also can exhibit similarities in the stressing dynamics pointing towards a common underlying mechanism for defect generation.

It is clear from Fig. 4 that a first order kinetics model ($recovery \propto \exp(-t/\tau)$) as that employed by Huang [6] for $t>100$ hours cannot describe the dynamics of the recovery for $t<10000$ s, as concluded by Huang. This kind of near *log t* behaviour has been previously reported for e.g. hole detrapping in MOS capacitors [16]. Detrapping of positive charge in the oxide has also been reported to be the recovery mechanism of stressed bipolar transistors subjected to a forward bias or RT anneal giving a *log t* recovery transient [11]. The change in oxide charge can also be observed by monitoring the change in flat band voltage, $\Delta V_{FB}(t)/\Delta V_{FB}(0)$, which is also included in Fig. 4. The data shown are for two different temperatures (28 and 103°C) for a MOS capacitor stressed using avalanche injection (from Lakshmanna and Vengurlekar [16]).

## 4. Conclusions

We have shown that the stress dynamics of bipolar transistors and ultra thin oxide MOS devices can

display similarities. In addition, the recovery dynamics and temperature dependence also show similarities suggesting that the defect creation and passivation mechanisms in bipolar transistors and MOS devices are related. This further indicates that we are dealing with the same defects in both cases, motivating further investigations in the possible correlation between degradation models for both bipolar transistors and MOS devices. An important issue is to model the impact of the defects on the current in bipolar transistors in order to re[illegible] dependence of the defect generation on curre[illegible]ectric field.

## References


[1] Joshi S. P., Lahri R. and Lage C. Poly emitter bipolar hot carrier effects in an advanced BICMOS technology *Proceedings of the IEDM* (1987) 182-5.

[2] McAndrew C. C., Kizilyalli I. C. and Bude J. D. BiCMOS $h_{fe}$ degradation: causes and circuit solution *Proceedings of the BCTM* (1994) 197-200.

[3] Collins D. R. Excess current generation due to reverse bias p-n junction stress *Applied Physics Letters.* 13 (1968) 264-6.

[4] Burnett J. D. and Hu C. Modeling hot-carrier effects in polysilicon emitter bipolar transistors *IEEE Transactions on Electron Devices* 35 (1988) 2238-44.

[5] Momose H. S. and Iwai H. Analysis of the temperature dependence of hot-carrier-induced degradation in bipolar transistors for Bi-CMOS *IEEE Transactions on Electron Devices* 41 (1994) 978-87.

[6] Huang C. J., Grotjohn T. A., Sun C. J., Reinhard D. K. and Yu C. C. Temperature dependence of hot-electron degradation in bipolar transistors *IEEE Transactions on Electron Devices* 40 (1993) 1669-74.

[7] Shimamoto H., Tanabe M., Onai T., Washio K. and Nakamura T. Test Structure and Experimental Analysis of Emitter- Base Reverse Voltage Stress Degradation in Self-Aligned Bipolar Transistors *IEICE Trans. Electron.* E-79-C (1996) 211-7.

[8] Neugroschel A., Sah C.-T. and Carroll M. S. Accelerated Reverse Emitter-Base Bias Stress Methodologies and Time-to-Failure Application *IEEE Electron Device Letters* 17 (1996) 112-4.

[9] Sun C. J., Reinhard D. K., Grotjohn T. A., Huang C. J. and Yu C. C. W. Hot-electron-induced degradation and post-stress recovery of bipolar transistor gain and noise characteristics *IEEE Transactions on Electron Devices* 39 (1992) 2178-80.

[10] Horiuchi T., Burnett J. D. and Hu C. Bipolar transistor degradation under dynamic hot carrier stress *Solid State Electronics* 38 (1995) 787-9.

[11] Neugroschel A., Sah C.-T., Carroll M. S. and Pfaff K. G. Base current relaxation transient in reverse emitter-base bias stressed silicon bipolar junction transistors *IEEE Transactions on Electron Devices* 44 (1997) 792-800.

[12] Maserjian J. In: *The physics and chemistry of SiO2 and the Si-SiO2 interface-2*, C. Helms and B. E. Deal, Eds. New York, 1988.

[13] Lundgren P., Andersson M. O. and Farmer K. R. Post-metallization annealing of metal-tunnel oxide-silicon diodes *Journal of Applied Physics* 74 (1993) 4780-2.

[14] Neugroschel A., Sah C.-T. and Carroll M. S. Degradation of Bipolar Transistor Current Gain by Hot Holes During Reverse Emitter-Base Bias Stress *IEEE Transactions on Electron Devices* 43 (1996) 1286-90.

[15] Ragnarsson L.-Å., Lundgren P. and Jauhiainen A. Electrical Reliability of Ultra Thin Remote Plasma Deposited Oxides on Silicon *Microelectronic Engineering* 36 (1997) 65-7.

[16] Lakshmanna V. and Vengurlekar A. S. Logarithmic detrapping response for holes injected into SiO2 and the influence of thermal activation and electric fields *Journal of Applied Physics* 63 (1988) 4548-54.

PERGAMON

Microelectronics Reliability 38 (1998) 1115–1119

MICROELECTRONICS RELIABILITY

# Electrical parameters degradation law of MOSFET during ageing

Ch. Mourrain, Ch. Tourniol, M J. Bouzid

*France Telecom CNET Grenoble, 28 ch. du vieux chêne, BP 98, 38243 Meylan Cedex France*

**Abstract**

A new method to extract the different electrical parameters lifetime of MOS transistors submitted to hot carriers degradation is proposed. This method leads to error on the lifetime below 15%, even if the parameter variation measurement reaches only 8%. The robustness of this method has been tested for various biases of stress and different technologies representative of different ageing mechanisms. Finally this method is a good indicator of the degradation modes occurring during the stress. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

The devices lifetime is one of the criterion used to qualify a technology. So it is extremely important to get an accurate value. The current laws used to describe the parameters evolution during the stress are not fully satisfactory. The purpose of this paper is to give a new way to fit this evolution and to get more information about the ageing mechanisms.

## 2. Experimental

The ageing stresses have been performed on W/L= 20µm/0.18µm NMOS and PMOS transistors. Three different architectures have been tested :

Table 1
Main differences in the process steps.

| Wafers number | Wafer 2 | Wafer 4 | Wafer 11 |
|---|---|---|---|
| PMOS channel | Ph 70Kev 1.2E13 | As140Kev 1.1E13 | As140Kev 4E12 |
| NMOS channel | B 30 Kev 1.2E13 | In 170 Kev 1.1E13 | In 170 Kev 5E12 |
| N well implantation | _ | _ | As 40Kev 4.5E12 |
| Gate material | Si | Si | $SiGe_{55\%}$ |
| Gate predoping | dual | dual | P+ |
| Spacer | PECVD | PECVD | PECVD |

Those wafers have been chosen because they are really different and lead to different degradation mode regards to hot carriers degradation.

## 3. Discussion and results

There are two well-known laws actually used to describe the parameters evolution during the stress : the power law at$^n$ [1,2,3] and the logarithmic law [2,4]. But both of them have their limitations.

The $\log(\Delta X/X_o)$ versus $\log(t_{stress})$ plot is used to extract the lifetime with the at$^n$ law. This should give a straight line. But because there is some attenuation in the parameter degradation the curve may have different shape as shown on figure 1.

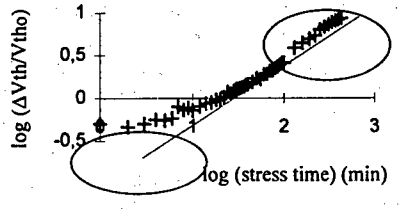



Fig. 1. Limitations of the at$^n$ law. Vth evolution versus log(stress time). NMOS, 10µm/0.18µm, wafer 4.

In the same time the sign of the threshold voltage variation may change during the stress. So the only points that can be used should have the same sign (because of the log-log law). This means that the law does not describe accurately the phenomenon and that an important information on the degradation mode change is occulted.

The $\Delta X/X_o$ versus $\log(t_{stress})$ plot is used to extract the lifetime with a first order log(t) polynomial law. But as shown on the figure 2, the ageing rate is not constant so the same straight line cannot describe all the domain. When the measurement is stopped before 10% of variation the

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00139-5

error made on the lifetime extrapolation can clearly be more than one order of magnitude.

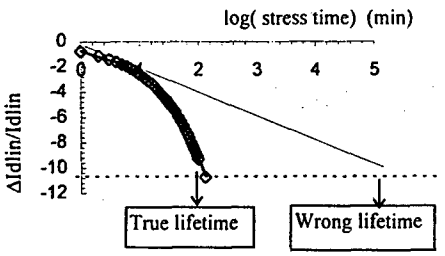



Fig. 2. Limitations of the first order polynomial law for Idlin variation versus log(stress time). NMOS, 10μm/0.18μm, wafer 2.

The purpose of this work was to find out a law which could both describe the attenuation in the degradation and some reverse in the sign variation. This law should also helps to understand the mechanisms involved during the stress.

### 3. 1. Determination of the degradation law

The lifetime has to be extracted even if the parameter variation sign changes during the stress. So the relative variation : ΔX/Xo will be the Y axe. The X axe chosen is $\log(t_{stress})$. So the plot is the same as in the figure 2.

Two laws have been tested to fit the experimental data : a second and third order polynomial. They have been tested on thirteen parameters : the linear drain current in direct (Idlin) and reverse mode (Idliv), the saturation drain current in direct (Idsat) and reverse mode (Idsav), the threshold voltage in direct (Vth) and reverse mode (Vthv), Mo (=μo*Cox*(W/L)), the two mobility attention factors ThetaG and Theta2, the maximal transconductance $Gm_{max}$, the access resistance, the sub-threshold slope S, the leakage current Idoff.

To estimate the fit quality on the parameter (X) variation versus the stress time, three criteria were monitored : the parameter variation after the first step of stress : $(\Delta X/Xo)_{\%1}$, the lifetime corresponding to 10% variation of the parameter or if not reached to the time corresponding to the last stress : $(\tau_X)$ and the general aspect of the curve monitored with the root mean square ($R^2$).

The accuracy criterion that we have fixed for both the relative error on $(\Delta X/Xo)_{\%1}$ and on the lifetime was extremely severe : 10%.

For all the study the stresses have been stopped when Idlin variation reaches 10%.

### 3. 2. Fit on all the measured points

In a first time, the fit has been performed on all the measured points in order to evaluate if the laws chosen were suitable to describe the parameter degradation.

The lifetime corresponding to 10% of various parameters have been extrapolated and compared to the measured time corresponding. For Idsat, Idsa2 and S, the 10% variation was not reached during the all measurement, so the extrapolated lifetime presented in the following corresponds to their actual variation percentage. The stress conditions (Vds, Vgs) correspond to the maximum substrate current and the lifetime related to Idlin was about 30mn. The results are gathered in the table 2 :

Table 2
Error between the ΔX/Xo @ log(t) 2nd and 3rd order polynomial laws and measurement for the percentage variation of the first step and for the lifetime. NMOS 10μm/0.18μm, wafer 11. All the measured points (i.e. ΔIdlin=10%) are used for the fit.

| X | error on $(\Delta X/X_o)$% 1 | | error on $\tau_X$ | |
|---|---|---|---|---|
| | 2nd order $\Delta x/x_o$@log(t) | 3rd order $\Delta x/x_o$@log(t) | 2nd order $\Delta x/x_o$@log(t) | 3rd order $\Delta x/x_o$@log(t) |
| Idlin | -0.44% | -0.2% | 2% | 2% |
| Idliv | -0.46% | -0.06% | +0.4% | +0.4% |
| Idsat | -4.44% | -0.9% | +5.33% | +3.33% |
| Idsav | -0.8% | -0.1% | -0.7% | -0.7% |
| S | +18.8% | +3.5% | +5.33% | +2.67% |
| Gm | +1.9% | -0.3% | -0.33% | +1.33% |

The error is always below 5% when choosing the proper degree for the polynomial law to fit the curve $\Delta X/X_o$ versus the logarithm of the stress time.

The error on the parameter variation during the first stress step ($(\Delta X/Xo)_{\%1}$) with the third order polynomial exhibits a better accuracy than with the second order. The figure 3 clearly show the reason : the maximum of the polynomial will not be on the first point but later on.

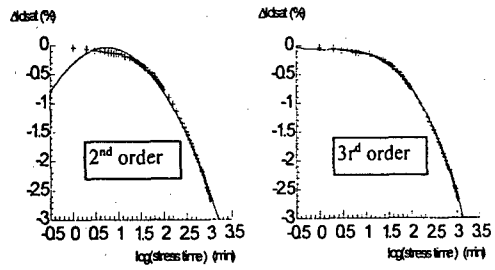



Fig. 3. 2nd & 3rd order polynomial law fit for Idsat variation versus log (t). NMOS 10μm/0.18μm, wafer 11.

The worse 2nd order fit appears for the parameters which are slightly degraded during the stress. Typically Idsat and S are the more concerned for all the technologies that have been tested. In the same way as far as the stress bias is reduced the third order law will become more accurate for such parameters as Idlin or Idsav. When the degradation rate of the parameter is high the 2nd order is more indicated. This will be pointed out later on.

So we demonstrate that the curve of the degradation of MOSFET's parameters : $\Delta X/X_o$ versus the logarithm of the stress time is perfectly described with either a 2d order or a 3rd order polynomial law.

### 3. 3. *Fit on 80% of the measured points*

In the next step the extrapolation have been performed, at first for one condition of stress, on the measured points corresponding to the first 8% degradation of Idlin. This means that 80% of the measured points will be used for the fit but because the stress has been performed until 10% of the parameter variation the actual and extrapolated lifetimes will be compared. The results on the lifetime prediction are gathered in the table 3 :

Table3
Error between the ΔX/Xo @ log(t) 2nd and 3rd order polynomial laws and measurement for the lifetime. Wafer 4, NMOS 10μm/0.18μm. 80% of the measured points (i.e. ΔIdlin=8%) are used for the fit.

| X | $\tau_X$ measured | Error on $\tau_X$ with 2nd order $\Delta X/X_o$@log(t) | Error on $\tau_X$ with 3rd order $\Delta X/X_o$@log(t) |
|---|---|---|---|
| Idlin | 130 min | 17% | +2.6% |
| Idliv | 130 min | +6.5% | +8.5% |
| Idsat | 130 min | +35.9% | +6.5% |
| Idsav | 45 min | -0.2% | -0.2% |
| S | 130 min | +70.3% | +21.5% |
| Gm | 16 min | -7.5% | -8.8% |

Remark : For Idsat, Idsa2 and S, the 10% variation was not reached during the all measurement, so the extrapolated lifetime presented in the following corresponds to their actual variation percentage.

The lifetime prediction for this stress bias is extremely accurate. The error is less than 9% as far as the correct order for the polynomial law is chosen. The 3rd polynomial law is more accurate for the parameters slightly degraded as observed when all the measured points were used for the fit.

Then the ageing has been performed at various stress bias to find out if there was a correlation between the order of the polynomial fit and the degradation rate. The stress biases have been chosen in order to get lifetime ranging from 10 minutes up to 1000 minutes. The results, corresponding to the first stress step percentage variation and to the lifetime of Idlin extrapolated when using the first 8% measured points compared to the measured values for all the stress conditions, are presented in the table 4 :

Table 4
Measurement-Extrapolation error on the first minute variation and lifetime of Idlin for different biases of stress. NMOS 10μm/0.18μm, wafer 11. 80% of the measured points (i.e. ΔIdlin=8%) are used for the fit.

| $\tau_{Idlin}$ measured | error on $(\Delta X/X_o)\%_t$ | | error on $\tau_x$ | |
|---|---|---|---|---|
| | 2nd order $\Delta x/x$@log(t) | 3rd order $\Delta x/x$@log(t) | 2nd order $\Delta x/x$@log(t) | 3rd order $\Delta x/x$@log(t) |
| 1270 min | +3.97% | +0.45% | -21% | -7% |
| 1150 min | +2.33% | +0.6% | +27.4% | +10.8% |
| 190 min | +2.26% | +1.63% | -30.1% | -13.35% |
| 95 min | +0.46% | +0.4% | -14.06% | -15.08% |
| 75 min | +0.21% | +0.06% | -9.96% | -6.3% |
| 45 min | +0.32% | +0.11% | -3.86% | -17.54% |
| 30 min | +0.29% | +0.12% | -0.49% | -20.95% |
| 26 min | +0.03% | +0.12% | +11.21 | +0.4% |
| 10 min | +0.2% | 0% | +5.25% | +94.5% |

This table clearly exhibits that for fast degradation (high biases of stress), the 2nd order of the polynomial law is totally correct but the 3rd order is not adapted. And for slow degradation (low biases of stress) the 3rd order is the best choice.
This is not in contradiction with the postulate that the degradation mode must be the same for all the stress bias to give a valid lifetime at the supply voltage. With this postulate the slope of log(Δx) versus log(t) must be independent of the stress bias. The measurement on the wafer 11 (see Fig. 4) exhibit that such behaviour is observed with the log-log scale.

The figure 5 shows the same results but in a lin-log scale, which is the format proposed in this paper to fit the degradation curves. The parameter variation is no more smoothed over. And it is clear that the order of the polynomial law cannot be the same for high and stress biases.

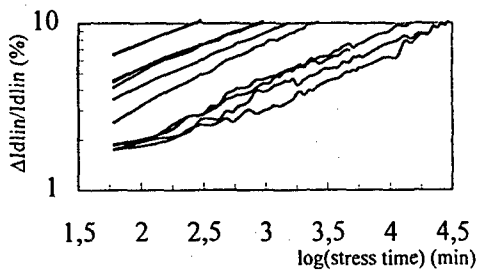



Fig. 4 Idlin variation versus log(stress time) for various stress biases, in log-log scale.

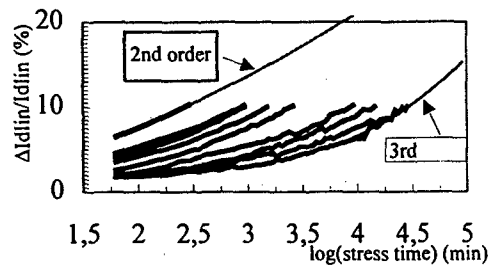



Fig. 5 Idlin variation versus log(stress time) for various stress biases, in lin-log scale.

The same study has been performed on all the parameters and strengthen those conclusions.

A more detailed study, for example on the Vth, interface state density (Nss) or Mo evolution, point out that the 2nd order will be used when the degradation rate is high. This is typically the case when the lifetime is lower than 100 minutes.

The 3rd order will be used when :

- The degradation rate is low. Typically when the stress bias leads to lifetime up to 100 minutes.
- Two degradation phenomena are in competition as positive charges created into the oxide and interface states charged negatively at the interface. This will induce a change in the sign of threshold variation.
- The drain saturation current is studied. In that case the variation percentage is extremely low as far as the degraded region is included in the pinch off region and increased quickly when it is beyond.

A good indicator to detect if the 2nd or 3rd order is need, is the error made on the variation during the first step of the stress between the measured value and the 2nd order extrapolated value. If this error is close to zero the second order is the best choice, if it is higher than a threshold value the third order will be recommended, as shown on figure 3.

If the third order of the polynomial law is used a second verification has to be done. Because with this mathematical solution a roll up can be observed and then no solution will be found as shown on figure 6.

But each time such behaviour has been observed, the lifetime error made with the second order law was below 10% and the third order was not needed.

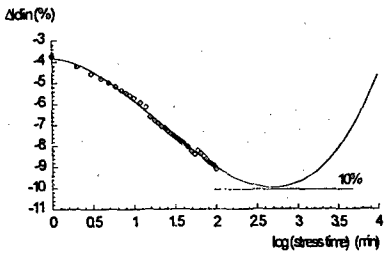



Fig. 6. Roll up with a 3rd order polynomial extrapolation. NMOS, 10μm/0.18μm, wafer 11.

An algorithm has been created and tested at the CNET Grenoble in order to get an automatic treatment of the extrapolation of the lifetime for all the parameters followed during ageing.

### 3. 4. Results

The results on the threshold voltage and Idlin evolution are presented respectively on figure 7 & 8.

As shown on the figure 7 the phenomena involved in the electrical characteristics degradation of the transistor are different. For the wafer 11 the competition between positive charges creation into the oxide and interface states creation charged negatively at the interface clearly appears. The sign of the threshold voltage variation during the stress changes. But this behaviour is correctly modelled with the third order of the ΔX/Xo versus log(t) polynomial law. And this allows us to get access on this important information about the change of the degradation mode dominance.

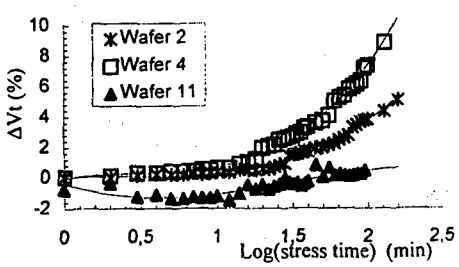



Fig. 7. Extrapolation on Vth using the CNET algorithm. NMOS, 10μm/0.18μm, all wafers.

The figure 8 exhibits that the second order polynomial law is perfectly adequate to model the linear drain current evolution during hot carriers

degradation for all the wafers even if the degradation mode is different.

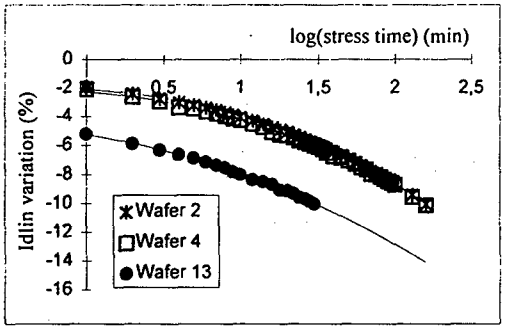



Fig. 8. Extrapolation on Idlin using the CNET algorithm. NMOS, 10μm/0.18μm, all wafers.

The lifetime has also been extracted with this algorithm when the measured points represent only 9% of the parameter variation. In such case 70% of the error made on the lifetime determination on all the electrical parameters were below 3%, 85% of the error were below 5% and 100% of the error were below 10%.

The study has not been performed for lower parameter variation than 8% because in such case the attenuation of the degradation rate is not reached for most of the parameter. Thus the fit will still be correct but non representative of the real behaviour of the device under stress.

On the other hand to stop the ageing when 9% variation is reached saves around 30% stressing time and about 50% for 8% variation. This is quite interesting to get accurate and fast results.

## 4. Conclusions

The electrical parameters degradation of MOS transistors submitted to hot carriers stresses can no more be fitted with only one equation as $at^n$ for example. This because the degradation rate will not be the same for all the parameters and for the different stress biases. Two laws are therefore necessary to describe all the cases.

A second and third order polynomial law have been proposed to fit the curve $\Delta X/Xo$ versus the logarithm of the stress time.

Using the appropriate equation the error on the lifetime was lower than 15% for 98% of the cases when the measured points used for the extraction correspond to 8% of degradation. 100% of the error is lower than 10% if the stress is stopped for 9% parameter variation.

The test duration saved when the stress stops for 9% of the parameter variation is around 30% and about 50% for 8% variation.

So using this extrapolation method the ageing measurement can be both faster and more accurate.

Finally the degree of the polynomial law is a good indicator of the degradation mechanisms of the MOS transistor.

## Acknowledgements

The authors wish to acknowledge the contributions of GRESSI clean-room, material and architecture teams in processing of samples used in this work. A part of the results have been generated in the frame work of the European Project ACE ESPRIT.

## References

[1] Takeda E, Yang C.Y, Miura-hamada A. Hot-Carrier Effects in MOS Devices. Academic Press, 1995.

[2] Bravaix A, Vuillaume D. Lifetime prediction methods for p-MOSFETs: A comparative study of standard and charge pumping lifetime criteria. IEEE Transaction on Electron devices, vol. 42, no. 1, January 1995

[3] SUN S.W, ORLOWSKI M, FU K. Parameter correlation and modelling of the power-law relationship in MOSFET Hot-carrier degradation. IEEE Electron Device Letters, Vol. 11, N° 7, July 1990.

[4] WANG Q, KRAUTSCHNEIDER W.H, WEBER W. Explanation and model for the logarithmic time dependence of P-MOSFET degradation. IEEE Electron Device Letters, Vol. 12, N° 5, Mai 1991.

PERGAMON

Microelectronics Reliability 38 (1998) 1121–1125

MICROELECTRONICS RELIABILITY

# Investigation Of The Intrinsic $SiO_2$ Area Dependence Using TDDB Testing And Model Integration Into The Design Process

**James Prendergast [1], Nigel Foley [1], John S. Suehle [2]**

(1) Analog Devices, Raheen Industrial Estate, Limerick, Ireland.
Ph: +353-61-229011, Fx: +353-61-301164
(2) National Institute Of Standards And Technology, Gaithersburg, Maryland, USA
Ph: +301-975-2247, Fx: +301-948-4081
Email: james.prendergast@analog.com, nigel.foley@analog.com, john.suehle@nist.gov

**Abstract:**

This paper models the area dependency for thin $SiO_2$ films (1.2E-7 to 1E-2 $cm^2$) using Time Dependent Dielectric Breakdown testing over a wide range of electric fields and test temperatures. The data generated indicates that the field and temperature acceleration factors are the same for all the areas tested indicating that the failure mechanism is the same even though the times to failure are different for all the area sizes. The paper will explain and model the area effect on TDDB lifetime and use the model to predict gate oxide reliability in the design cycle. © 1998 Elsevier Science Ltd. All rights reserved.

## Introduction:

To date there has been very little comprehensive work done on the area dependence of capacitors using Time Dependent Dielectric Breakdown testing. The area dependence was initially investigated by Sune et al. [1] and indicated that it existed for Qbd and potentially for TDDB. Recent work at IRPS'97 tutorials by El-Kareh and Tonti [2] and Degraeve [3] also indicated the same area dependence but these investigations were limited by either the range of areas investigated or the range of test conditions used.

This paper provides a thorough investigation into the oxide area dependence over 5 orders of magnitude using multiple temperatures and electric fields in order to understand the breakdown mechanism, failure statistics, and model the area dependency. The paper will outline the structures tested, and the thermal and field acceleration factors generated for the different area sizes. It will also outline the reason for the area dependence and indicate how existing models must be modified to account for this dependency when predicting product reliability. The paper will also discuss the model implications and why it is necessary to consider the range of oxide areas used in a product when predicting oxide reliability.

## Test Structures:

The area analysis was conducted on flat P-type capacitors and on some NMOS FET structures to increase the area spread being investigated. The capacitors and transistors were fabricated on a 0.6 μm dual poly dual metal CMOS process with a target deposited oxide thickness of 125 Angstroms. The process had been extensively characterized and monitored since release and all monitored lots showed only an intrinsic distribution.

The structures included a 20 x 0.6 μm transistor, the minimum geometry allowed on the process. All the structures tested were rectangular in shape. Testing was performed in accumulation and a full listing is given in Table 1. below.

| Area $cm^2$ | Structure | Area $cm^2$ | Structure |
|---|---|---|---|
| 1.2 E-7 | NMOS | 0.001 | P Cap |
| 1 E -5 | P Cap | 0.0025 | P Cap |
| 4 E -5 | NMOS | 0.005 | P Cap |
| 0.0001 | P Cap | 0.01 | P Cap |
| 0.0005 | P Cap | | |

Table 1
Areas Used And Structure Types

The sample size used in each temperature/voltage test combination was 16 units and the structures were packaged in 16 lead side brazed ceramic packages using aluminum bond wires. This sample size was deemed sufficient as extensive characterization and monitor data generated since the process was released indicated that the oxide behaved in a purely intrinsic fashion.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00140-1

## Test Equipment:

The test equipment used to conduct the investigation was commercially available. It consisted of a mainframe which housed voltage and ammeter modules. These supplied the specified voltage to the units under test and ammeters measured the current on a continuous basis once the tests started. The devices under test were placed in ovens with very tight temperature control using multi-layer printed circuit boards. The test software operated in a MS Windows environment and the system was microprocessor controlled.

## Failure Distributions:

Figure 1 is an indication of the distributions seen during the evaluations. The sample size used in all cases was 16 units per temperature and electric field combination as the extensive monitor data we had did not indicate an extrinsic population.

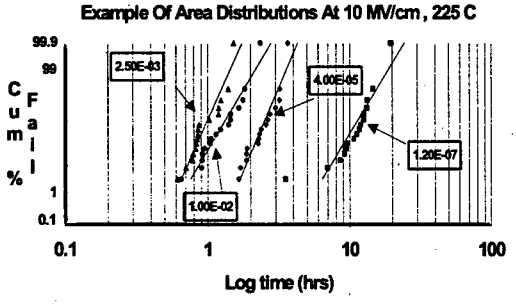



Figure 1
Example Of The 10 MV/cm, 225 C Failure Distributions

The initial data set covering all the areas shown in figure 1 above was generated at 10 MV/cm and 225 degree C. This figure indicates an area dependence and the smaller the capacitor area, the longer the time to failure.

## Area Dependence:

The data set shown in figure 1 above was expanded to include all areas and is shown in figure 2.

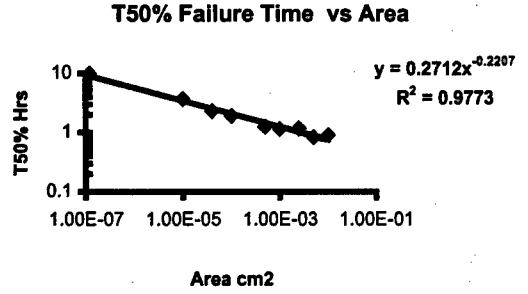



Figure 2
Failure Time (Hrs.) vs Area At 10 MV/cm and 225 C.

This figure shows the log of the 50% cumulative failure time plotted as a function of log area ($cm^2$). The equation in figure 2 gives the slope of the line followed by the correlation coefficient indicating the potential to model the area dependence. Based on this result, further tests were carried out at other electric fields and temperatures to investigate the area dependence over different electric fields and temperatures.

These additional tests allow the calculation of the thermal and field acceleration parameters for the different area sizes verifying that the failure mechanisms are the same for all area sizes. The results are shown in figure 3.

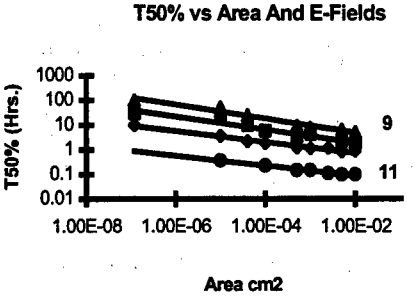



Figure 3
Failure Times vs. Areas Tested For Various E Fields

Figure 3 shows the data sets generated at 9, 9.5, 10 and 11 MV/cm (denoted by numbers) and 225 degree C for the range of areas shown in Table 1. The lines are parallel for all the electric fields tested, indicating that the behavior is the same for the range of E fields tested.

## Acceleration Factors:

The next step to validate the above assumption regarding the failure mechanism was the calculation of the thermal activation energies for the various area sizes being investigated.

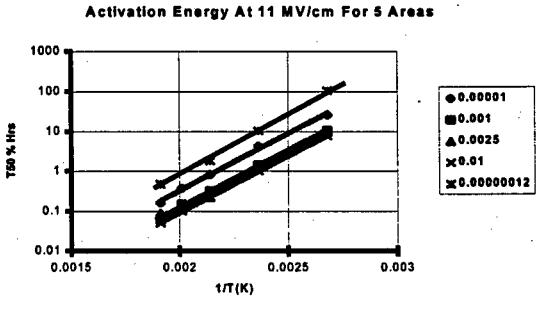



Figure 4
Thermal Activation Energies

The curves in Fig. 4 were generated at 10 and 11 MV/cm for a cross section of areas as shown in figure 4. The absolute value of the slopes is related to the activation energy

(Ea), and these are the same for all areas investigated as shown in Table 2 below.

| Area $cm^2$ | 0.01 | 0.001 | 0.0025 | 0.00001 | 1.2E-7 |
|---|---|---|---|---|---|
| Ea@10MV/cm | 0.76 | 0.72 | 0.73 | 0.75 | |
| Ea@11MV/cm | 0.55 | 0.52 | 0.52 | 0.53 | 0.55 |

Table 2
Activation Energies At 10 and 11 MV/cm

The behavior and results are consistent with previously reported data [4,5] and indicate an increasing activation energy with a decreasing electric field. These results support the fact that the failure mechanism behaves the same for all temperatures and is consistent with previously reported data.

The field acceleration factor was next calculated and is shown in figure 5 for a cross section of the areas examined, but it was calculated for all areas. The graph indicates that the slopes of the lines for each area are the same, indicating that the field acceleration is constant with area size. This was confirmed by calculating the field acceleration factor and plotting the calculated value versus the area as shown in figure 6 and is approximately equal to 1 for all areas. The standard deviations (sigma values) of the individual distributions for each test condition and area (i.e. 10 MV/cm, 225 C and 0.01 $cm^2$) were analyzed. It is also noted that this analysis shown in figure 7 did not show any trends with area.

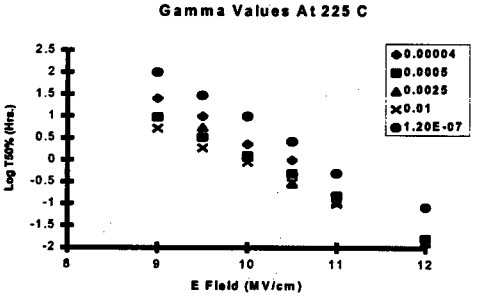



Figure 5
Field Acceleration Factors At 225 C

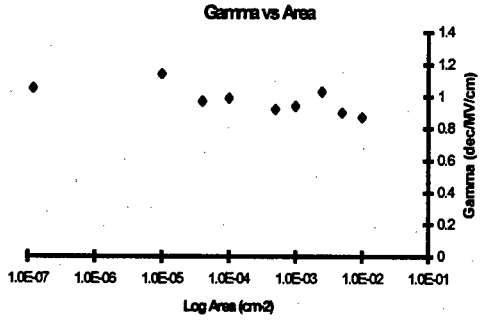



Figure 6
Gamma Values vs. Area

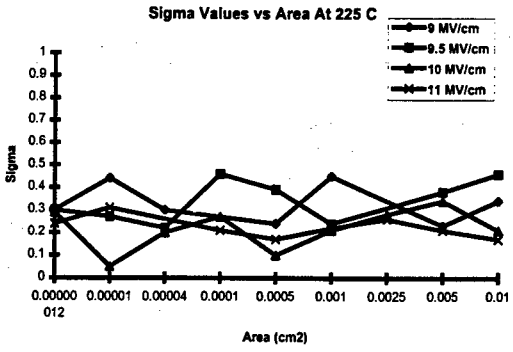



Figure 7
Sigma Values vs Area

The electric field dependence also indicates that the oxide follows the E model for all the areas examined. The acceleration parameters are the same for all areas, indicating that the final failure mechanism is the same for all areas. The major difference is in the times to failure for the different areas tested.

## Failure Mechanism:

In order to investigate the area dependency, the oxides were stressed under a hypervision system operating in a continuous mode. The output was displayed on a monitor and recorded on video, and provided a physical explanation for the different failure times seen on the different areas. The reason for the area dependency is related to the probability of failure which can occur and is best described in figure 8, which shows 3 different size area capacitors. The breakdown is not an instantaneous occurrence [7,8]. It is made up of many soft reversible intrinsic type breakdowns which occur before the final irreversible intrinsic breakdown occurs. The probability of the final breakdown occurring depends on the area of the capacitor as the number of soft breakdowns which occur is a function of the area of the capacitor. The larger areas have more soft breakdowns with a higher probability of one of these causing a final irreversible breakdown resulting in a lower failure time. In figure 8 below, three different area capacitors are represented by the rectangles. The soft breakdowns are indicated by an "s" and final one by an "f" and it is easy to appreciate that the number of soft reversible breakdowns that occur depends on the area being tested.

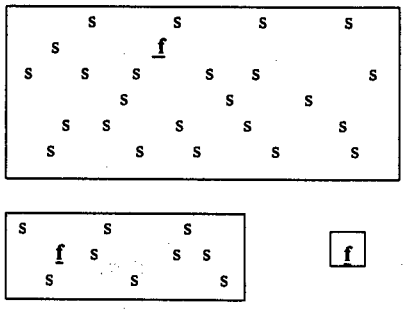



Figure 8
Example Of Failures Occurring

## Model Modification:

The generation of the above data indicates that the model which is applied to predict oxide reliability now needs to be modified. The data above indicates that to predict product lifetime, the area of the individual transistors within the product must be considered. From figure 2 above we can see that the time to failure, T50%, is dependent on the area at 10 MV/cm. This holds true for all electric fields tested and the equations generated for these electric fields are shown in table 3.

| E Field (MV/cm) | $TTF_t$ (Hrs.) |
|---|---|
| 9 MV/cm | $TTF_t = 1.3632\ A^{-0.2826}$ |
| 9.5 MV/cm | $TTF_t = 0.5584\ A^{-0.2629}$ |
| 10 MV/cm | $TTF_t = 0.2712\ A^{-0.2207}$ |
| 11 MV/cm | $TTF_t = 0.0358\ A^{-0.2001}$ |

A = Area in $cm^2$

Table 3
Life Time Equations vs Area And E Field

The coefficients and exponents are plotted in figures 9 and 10. These graphs indicate that the coefficients and exponents in the above equations can be plotted versus the electric field with a good correlation. Figure 9 indicates that coefficient is an exponential function of the electric field while figure 10 indicates that the exponent is a linear function of the electric field for the range of electric fields tested.

The linear E model [6] (see equ. 1) is

$$TTFu = Exp\ -Ea/k(1/Tu - 1/Tt)\ Exp\ \gamma(Et - Eu)\ TTFt \quad (1)$$

where Eu and Tu are the use electric field and temperature
Et and Tt are the test electric field and temperature
and TTFt is the time to failure at the test conditions.

The TTFt is generally generated on a specific size test structure and applied to the circuit as a whole when doing lifetime predictions. Depending of the test structure size tested and the transistor areas used in the product this can be either an optimistic or pessimistic prediction. If the product uses a significant number of transistors which have an area greater than the test structure then the prediction is optimistic and vice versa. As a result the area dependence needs to be included in the prediction model. This is done by replacing the constant TTF term with a time to failure term (FT) which is dependent on individual transistor areas (see equ. 2).

$$TTFu = Exp\ -Ea/k(1/Tu - 1/Tt)\ Exp\ \gamma(Et - Eu)\ FT \quad (2)$$

where FT is defined as

$$C \times A^{-B} \quad (3)$$

in Hrs. where C and B are defined by equations 4 and 5 below

In the above equation A is the area of the capacitor in $cm^2$ while C and B from figures 9 and 10 are functions of the test electric field as follows

$$C = 0.0421\ Et - 0.6573 \quad (4)$$

and $$B = 2E\text{-}7\ Exp\ -1.8201\ Et \quad (5)$$

This makes the linear E model more complex and as a result the lifetime prediction for the gate oxide becomes more difficult to calculate. It is no longer feasible to do the prediction based on single area test structure data and the oxide transistor area needs must be included. In order for this to be effective the model must be included into the product design phase and reliability predictions must be performed during the design phase to ensure that the reliability requirements of the product are being met.

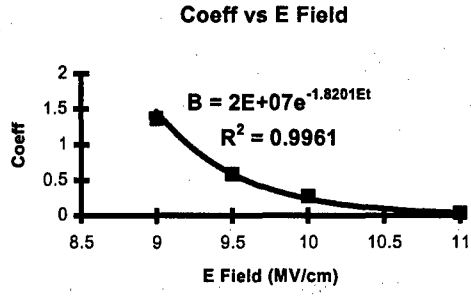



Figure 9
Coefficient Of TTF Equs. vs E Field

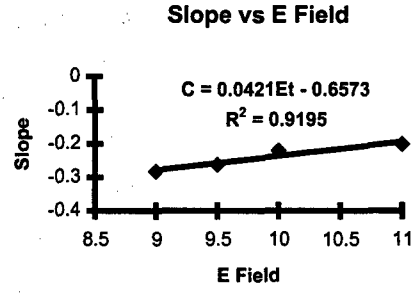



Figure 10
Slope Of TTF Equs. Vs E Field

Before product tapes are shipped, to make the mask sets used to generate the silicon for the product, a complete simulation is done. This simulation involves running the product at the proposed use conditions and collecting the gate voltages at various time steps during the simulation. The shorter the time steps the more accurate the data collection and as a result the more accurate the final lifetime prediction. The data collected at each time step during the simulation includes the transistor size, location and the gate voltage, which can easily be translated into an electric field, and the maximum voltage allowed by the design rules. This data for the full simulation is then stored in a file and is processed once the simulation is complete. The model parameters such as activation energies, field acceleration factors and test conditions are also stored in the file with the only user input being the maximum use temperature allowed for the product. The data collected from the simulation and the post processing allows the following:

A) Prediction of reliability at use/simulation conditions
B) Prediction of reliability at the maximum allowed process voltages.

Table4 below shows the results of applying the model to product in the design phase. From 70 C to 160 C the product continued to meet its reliability targets of 0.1% cumulative failure in 10 years. The data shown here was generated at 161 C indicating that the product would meet its reliability requirements as the product was specified up to 150 C. As the temperature increased from 161 C other smaller area transistors began to fail.

| Time (Yrs.) | Length (μm) | Width (μm) | Transistor |
|---|---|---|---|
| 9.6 | 0.6 | 100.8 | <U138.I1>M0 |
| 9.6 | 0.6 | 100.8 | <U4.I1>M0 |
| 9.6 | 0.6 | 100.8 | <U5.I1>M0 |

Table 4
Example Of Life Time Prediction On Real Product
Using The Maximum Allowed Process Voltages At 161 C.

The data generated in Table 4 above was generated for the worst case process conditions. The data collection, post processing the results and applying the model was in this case independent of the simulation results and took approximately 3 minutes to do for 66,000 transistors at 70 C with subsequent runs taking approximately 30 seconds each. Using the maximum transistor voltages gained during the simulation the results in this case indicated greater levels of reliability and transistors did not reach the maximum process allowable gate voltages.

## Summary:

This work reports on the area dependency for time dependent dielectric breakdown testing of thin $SiO_2$ films. The data derived the failure statistics for areas over 5 orders of magnitude indicates that smaller areas are more reliable. The paper outlined the model and indicates that the existing model must be modified to account for this area dependence when predicting product reliability. The data indicated the importance of the area when predicting oxide reliability and this must be comprehended in the design flow in much the same manner as the width and length dependency for electromigration. The paper also outlines a methodology for implementing the model into the design phase and shows results on a recently released product.

## Acknowledgments:

We would like to acknowledge the help and assistance of John Mac Namara (ADI) in the failure analysis group for his help and assistance with the Hypervision and video taping the failure mechanisms on the different area sizes. Also the contribution of David Lucas (ADI) must be acknowledged for providing the software which helped with the voltage extraction from the product simulations.

## References:

[1] On The Breakdown Of Very Thin SiO2 Films by Sune et al., Thin Solid Films, 185 (1990) Pgs. 347-362.
[2] Tutorial On Chip Reliability by El-Kareh and Tonti, IRPS 1997 Tutorial Session 1.
[3] Tutorial On Dielectric Breakdown by Degraeve, IRPS 1997 Session 7.
[4] Field And Temperature Acceleration Effects of Time Dependent Dielectric Breakdown in Intrinsic Thin $SiO_2$ by Suehle et al., IRPS 94, Pg 120.
[5] TDDB Characterization of Thin $SiO_2$ Films with Bimodal Failure Populations by Prendergast et al., IRPS 95 Pg 124.
[6] Acceleration Factors for Thin Gate Oxide Stressing J. W. McPherson and D. A. Baglee, Proc. Int. Rel. Physics Symp., Vol. 23, p.1, 1985.
[7] Investigation Of The Intrinsic $SiO_2$ Area Dependency Using TDDB Testing by Prendergast et al. IRW 1997, Pg 22
[8] The Non-Uniqueness of Breakdown Distributions in Silicon Oxides by Jackson et al., IRW 1997, Pg. 50

PERGAMON

Microelectronics Reliability 38 (1998) 1127–1131

MICROELECTRONICS RELIABILITY

# Two-step stress method for the dynamic testing of very thin (8 nm) $SiO_2$ films

R.Rodríguez, E. Miranda, M. Nafría, J. Suñé and X. Aymerich

*Dept. Enginyeria Electrònica. Universitat Autònoma de Barcelona. 08193 Bellaterra, Barcelona.Spain*

## Abstract

This work examines the use of a two-step electric test method which consists in applying a dynamic pre-stress followed by a static final stress (the same for the different pre-stresses) until breakdown. The method decreases the testing times under dynamic stress conditions and allows the comparison of the degradation introduced in the oxide for different types of pre-stress. Although the time-to-breakdown of the dynamically pre-stressed oxides measured during the final DC test is larger than the one measured in virgin oxides, the degradation is found to be a cumulative process.
© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

The dielectric breakdown of silicon dioxide, one of the most important failure mechanisms of ULSI ICs, has been traditionally studied under static (constant current or constant voltage) or ramped stresses. These tests are far from the actual operation of devices in ICs, since generally time-varying voltages will be applied to their terminals. In fact, when thin oxides are subjected to high-frequency bipolar stresses, a tremendous increase of the time-to-breakdown ($t_{BD}$) is observed [1-3]. Thus, dynamic stress conditions should be used for a more realistic oxide reliability assessment. However, due to the observed increase of $t_{BD}$ , the testing times become practically unaffordable unless the stress electric fields [1,2] or the temperature [3] are increased. However, this solution is not suitable because different oxide failure modes, which are inactive (or unimportant) under conditions closer to the IC operation, could be triggered [4].

In this work, we propose the use of a two-step procedure to reduce the testing times, applying a dynamic pre-stress followed by a static stress which finally induces the oxide breakdown. This method is based on the assumption that the breakdown is related to the degradation of the oxide during the stress and that the degradation is a cumulative process.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00142-5

## 2. Experimental procedure

The test structures used in this work were poly-Silicon gate MOS capacitors with ≈ 8nm thick oxide, n-type substrate (to allow injection from both the gate and substrate interfaces) and 3.24 $10^{-4}$ $cm^2$ area. The samples were stressed until breakdown or pre-stressed for a fixed period of time using static and dynamic voltage waveforms. A constant voltage of +10.6V (injection from the substrate) was chosen as the static stress condition and bipolar square voltages of 10.6V amplitude and frequencies of 10, 50 and 100Hz were the dynamic waveforms used.

The capacitors were pre-stressed using the waveforms described above for 15s. The pre-stresses were followed by a constant-voltage of +10.3V which finally led the oxide to the dielectric breakdown. This stress condition was the same for all the pre-stresses so that a comparison of the results is meaningful. Constant-voltage stresses of +10.3V on virgin oxides were also performed for comparison. The I-V characteristics (for both polarities) before and after the pre-stress and the I-t evolution during the final DC stress were also recorded. On the other hand, the time-to-breakdown of virgin oxides (without pre-stress) were measured to allow the comparison of the results obtained using the conventional test procedure with those of the two-step stress test proposed in this work.

## 3. Results and discussion

The statistical breakdown distributions of virgin oxides subjected to static and dynamic stresses until breakdown have been obtained. The mean values of the distributions are shown in table I. An increase of the mean time-to-breakdown is measured in the dynamically stressed oxides which grows with the stress frequency. In the analysed range of frequencies, $t_{BD}$ increases a factor of ≈20.

| | $t_{BD}$(s) |
|---|---|
| DC | 46 |
| 10Hz | 288 |
| 50Hz | 474 |
| 100Hz | 1007 |

Table I.- Mean time-to-breakdown measured in virgin capacitors during a +10.6V DC test and 3 bipolar stresses of 10.6V amplitude and different frequencies.

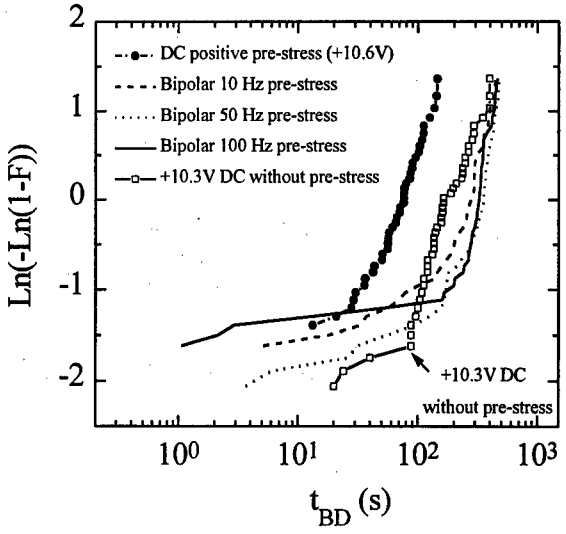



Figure 1.- Breakdown distributions as a function of $t_{BD}$ for different pre-stress conditions. $t_{BD}$ is in this case the time-to-breakdown measured during the final DC test.

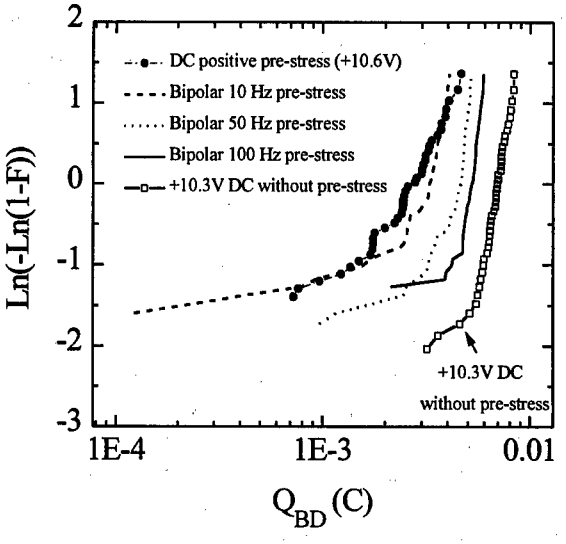



Figure 2.- Gumbel plot of the distributions in figure 1 as a function of $Q_{BD}$.

Figure 1 shows the Gumbel plot of the breakdown distributions obtained in the pre-

stressed oxides for the different pre-stress conditions. $t_{BD}$ refers to the time-to-breakdown of the pre-stressed oxides measured during the final +10.3V test. The breakdown distribution of the virgin oxide (without the application of pre-stress) is also shown for comparison. The samples that experienced breakdown during the pre-stress are considered by including them at $t_{BD}$=0s. The DC pre-stressed oxides show lower $t_{BD}$ than virgin oxides, but in the case of the dynamically pre-stressed capacitors, the measured $t_{BD}$ is larger. In this way, an increase of reliability is observed for the dynamically pre-stressed oxides (as also observed for other pre-stress conditions [5]) and one could conclude that the degradation of the oxide is not a cumulative process. However, if the injected charge to breakdown ($Q_{BD}$), a better suited magnitude to characterise the oxide breakdown, is used as the breakdown variable instead of $t_{BD}$ we realise that this conclusion is not correct. One should notice that in conventional tests (one-step stresses) $Q_{BD}$ cannot be easily defined for the dynamic tests, since in bipolar tests it is rather difficult to take into account the effects of polarity reversal of the stress on the calculation of the injected charge. However, since in the two-step test procedure the final breakdown is produced by applying a constant stress, the current-time characteristic during this stress can be used to calculate by numerical integration the injected charge needed to provoke breakdown. Figure 2 shows the Gumbel plot of the breakdown distributions in figure 1 as a function of $Q_{BD}$. We can observe that the injected charge to breakdown for the pre-stressed oxides is always smaller than that found for virgin capacitors. Therefore, this plot shows that the degradation is a cumulative process: the pre-stress induces some degradation in the oxide and, consequently, the injected charge to breakdown measured during the final DC stress is smaller than in the virgin oxide.

To clarify the apparently contradictory results in figures 1 and 2, the evolution of the current recorded during the final DC stress has been analysed. Figure 3 shows this evolution for the different pre-stress conditions (the one of the unstressed capacitors is also included). All the evolutions converge as the stress time increases and only differences are observed at the beginning of the test [5]: for all the pre-stresses, the current level is lower than in the virgin oxide. The initial behaviour of the current in a constant-voltage test has been attributed to trapping/detrapping processes in native traps and to the fast generation of new trapping sites by the injected electrons. Therefore, the differences observed in figure 3 should be related to differences in these two processes during the pre-stresses. To

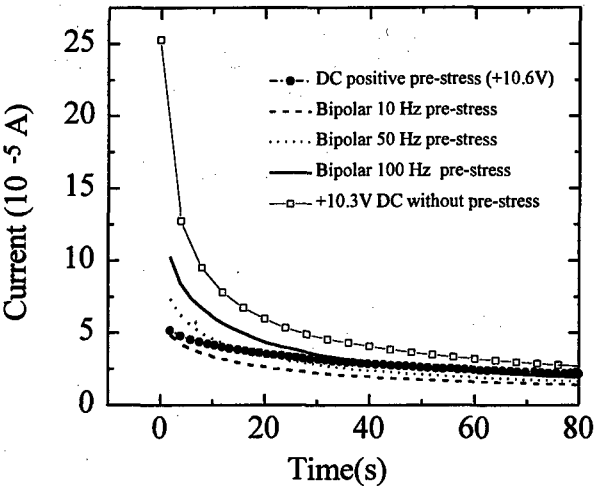



Figure 3.- I-t characteristics measured during the final DC stress for different pre-stress conditions.

confirm this assumption, the trapped charge distributions in the oxide after the pre-stress have been obtained from the I-V characteristics measured before and after the pre-stresses and are shown in figure 4. These distributions have been assumed to be indicative of the degradation level of the oxide (charge can be trapped in the defects that have been generated during the stress) [6]. It can be observed that the higher the stress frequency, the smaller the trapped charge density in the oxide (the lower the number of generated defects) and the closer to the centre of the oxide the centroid is. These results have been related to the polarity reversal in the dynamic stresses, which influences the oxide degradation process (the degradation introduced in the oxide decreases with the stress frequency [2]) and modifies the trapped-electron distribution [6]. Therefore, the pre-stresses lead to different trapped charge distributions in the oxide and, consequently, to different electric field profiles (electrons trapped near the cathodic interface will reduce the electric field and, consequently, less electrons will be injected). The injected current density will therefore depend on the pre-stress applied and, consequently,

the degradation rate during the final DC stress. This is the reason for the enhancement in the time-to-breakdown observed for dynamically pre-stressed oxides (figure 1). However, when $Q_{BD}$ is used as the breakdown variable, the modified electric field profile is already taken into account when the results are normalised to the number of injected carriers (charge) and no reliability improvement is observed. When $Q_{BD}$ is used, a cumulative degradation is always observed in the studied oxides (see figure 2) and the use of two-step stresses is therefore feasible.

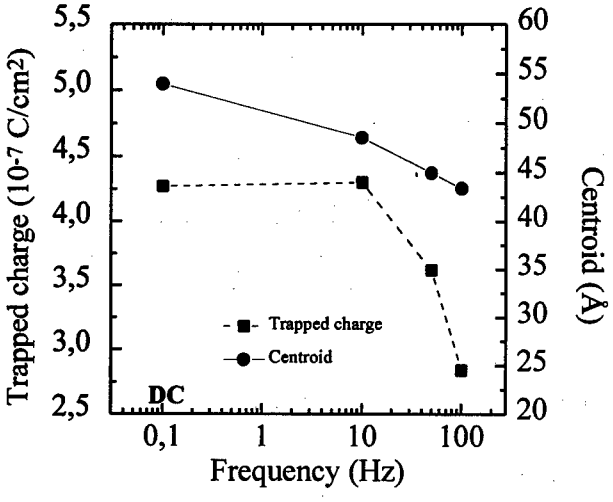



Figure 4.- Mean values of the average trapped charge density (negative) and centroid (measured from the poly-Si/$SiO_2$ interface) after the positive DC and bipolar pre-stresses

Once the feasibility of the two-step tests has been proved, we must verify the coherence of the data in the Gumbel plot as a function of $Q_{BD}$ (figure 2) and the results obtained in the conventional tests ($t_{BD}$ in table I) . Figure 2 shows that for the pre-stressed oxides $Q_{BD}$ increases with the stress frequency, being the lowest $Q_{BD}$ the one corresponding to the DC test and the highest to the 100Hz stress. This result agrees with the $t_{BD}$ data shown in table I for one-step tests: the degradation induced in the oxide by the dynamic tests decreases as the stress frequency increases [2], so that $t_{BD}$ increases in one-step tests and $Q_{BD}$ does so in two-step tests. Moreover, the distributions are ordered in frequency, i.e., the higher the pre-stress frequency the higher the $Q_{BD}$ , in agreement with the $t_{BD}$ data.

The use of $Q_{BD}$ as the breakdown variable also allows the comparison of the data in terms of the degradation process. When using this plot, information about this process can be obtained: the slope of the intrinsic part of the Gumbel plot is indicative of the critical number of defects that should be generated to provoke the oxide breakdown [4,7]. The smaller the slope, the lower number of defects are needed to produce breakdown. Figure 2 shows that the slope of the intrinsic zone of the distribution is clearly smaller in the DC pre-stressed oxides. This indicates that the degradation induced in the oxide during this pre-stress is greater. That is, a large amount of defects have been generated during the static pre-stress, so that the number of defects that has to be generated to provoke breakdown during the final DC test is lower and, hence, the breakdown would be induced earlier, as the breakdown data indicates. Moreover, the values of the slopes of the breakdown distributions seem to be grouped. The slope of the DC and 10Hz distributions are similar and the same happens with the 50 and 100Hz stresses. The slope is larger in the second group which indicates that the degradation process is somehow different for the highest-frequencies [1,2], in agreement with the enhanced reliability observed in this case. Finally, the number of extrinsic breakdowns increases when the stress frequency decreases. That is, the number of defects generated during the pre-stress (that will be accounted for as extrinsic defects during the final DC test) is larger for the lowest frequencies or, in other words, the degradation that they have induced is larger.

Note in figure 2 the high sensitivity of the method. Although in the 100Hz case the samples are pre-stressed for only approximately one hundredth of their time-to-breakdown, the shift from the virgin distribution is easily measurable. This observation seems to indicate that the degradation of the oxide is mainly induced at the beginning of the test. To verify this fact, the oxides have been pre-stressed with +10.6V DC and 10Hz bipolar voltages for different pre-stress times and have been broken down with the +10.3V constant-voltage test. As expected for a cumulative degradation process, it has been observed that $Q_{BD}$ decreases as the pre-stress time increases. However, there is a limit pre-stress time (which depends on the pre-stress condition) for which $Q_{BD}$ does not decrease when

the pre-stress duration is further increased. This provides additional support to the idea that the degradation of the oxide mainly occurs at the beginning of the stress test. The final breakdown, however, can occur after a stress time which is orders of magnitude larger because the degradation rate rapidly decreases with time. Although the relationship between degradation and breakdown is essential, these can be considered as two stages which occur at different time scales. In this way, the use of the proposed method is physically meaningful: the pre-stress induces the degradation and the DC test provokes the final breakdown.

## 4. Conclusions

Similar information can be obtained from one or two-step tests when $Q_{BD}$ is the breakdown variable. Both kinds of test have shown that the oxide reliability increases when it is dynamically stressed, specially at the highest frequencies. However, the two-step test has two main advantages. On the one hand, the degradation of the oxide has been observed to take place mainly at the beginning of the test so that the testing times can be kept reasonably low by inducing the degradation during the pre-stress and causing the final breakdown with a constant-voltage test (the worst stressing condition). On the other hand, and more important, the oxide degradation can be induced with stress conditions closer to the actual operation of devices in ICs (dynamic or low-field). From all the above, we can conclude that the two-step test procedure can be a promising tool to evaluate the oxide reliability.

## Acknowledgements

This work has been partially supported by the Ministerio de Educación y Cultura under project number PB96-1162. The authors are grateful to the Centre Nacional de Microelectrònica (Barcelona) for sample preparation. R. Rodríguez and M. Nafría have benefited from the Access to Large-Scale Facilities Activity of the European Community's Training and Mobility of Researchers (TMR) Programme by gaining access to the Large Scale Facility in Microelectronics at the National Microelectronics Research Centre (NMRC) at Cork (Ireland) to perform the dynamic measurements. They are grateful to T. O'Shea of NMRC for technical support.

## References

[1] E. Rosenbaum, Z. Liu and C. Hu, IEEE Trans. Elect. Dev. Vol. 40, No. 12, p.2287 (1993)
[2] M. Nafría, J. Suñé, D. Yélamos and X. Aymerich, IEEE Trans. Elect. Dev., Vol. 43, No 12, p.2215 (1996)
[3] J.S. Suehle and P. Chaparala, IEEE Trans. Elect. Dev., Vol. 44, No. 5, p. 801 (1997)
[4] M. Nafría, J. Suñé and X. Aymerich, IEEE Trans. Elect. Dev., Vol. 40, No. 9, p.1662 (1993)
[5] A. Martin, P. O'Sullivan, A. Mathewson, J.S. Suehle, P. Chaparala, Solid-State Electron., 41(7), p.1013 (1997)
[6] R. Rodríguez, M. Nafría, J. Suñé and X. Aymerich, IEEE Trans. Electron Dev., Vol. 45, No. 4, pp. 881-888 (1998)
[7] M. Depas and M.M. Heyns, Microelect. Eng., Vol 36, p.21 (1997)

PERGAMON

Microelectronics Reliability 38 (1998) 1133–1138

MICROELECTRONICS RELIABILITY

# Temperature dependence of snap-back breakdown up to 300°C analyzed using circuit level model and simulation

Dirk Uffmann

*Laboratorium für Informationstechnologie, Universität Hannover, Schneiderberg 32, D-30167 Hannover, Germany, Tel. +49 511 762 5016, Fax +49 511 762 5824, e-mail uffmann@lfi.uni-hannover.de*

**Abstract**

*The MOS snap-back breakdown and its temperature dependence were investigated up to 300°C using silicided LDD-NMOS transistors. The snap-back sustaining voltage increases from 8.25V at room temperature to 8.9V at 300°C (for $L_{eff}$=0.56μm). By using extracted parameters for a simple lumped element model we explain this behaviour originating from an increasing avalanche breakdown voltage and slope of avalanche multiplication factor compensating the increase in bipolar gain with temperature.*

© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Numerous applications would benefit from the availability of temperature resistant electronics. For economical reasons the use of CMOS circuits is advantageous due to low cost and short term availability. Work has been performed to develop a digital CMOS ASIC cell library and analog building blocks for temperature resistant electronics [1,2].

The operation of CMOS at high temperatures leads to a variety of problems. Besides leakage currents and latch-up [3] another potential source of failure is the avalanche induced breakdown of the parasitic bipolar transistor, which is formed between source, drain, and bulk of the MOS transistor, the so-called snap-back. According to Dreike et al. [4] increased leakage at high temperatures can result in snap-back. Snap-back itself can be a precursor to latch-up (Polgreen and Chatterjee [5]). The snap-back behavior has been widely studied since the early works by Troutman [6], Toyabe et al. [7] and Sun et al. [8]. Snap-back has been considered either as a parasitic effect leading to failure of output transistors in the event of ESD [9,5], or as a useful mechanism in ESD input protection circuits [10,11]. In both cases considerable self heating of the devices may occur. Gaston et al. [12] investigated the temperature sensitivity of snap-back in the range from 25°C to 150°C. They found no significant change of snap-back sustaining voltage with temperature and postulated a compensation of the different temperature effects. To our knowledge there is no publication on the temperature dependence of snap-back up to 300°C. The basic influencing parameters such as leakage currents, substrate resistance, current gain of the parasitic bipolar transistor and avalanche breakdown voltage are certainly temperature dependent.

There has also been effort to model the breakdown behaviour of bipolar and MOS transistors on circuit level to allow implementation within CAD tools [13,14,15], even including electro-thermal effects [16]. However, work on parameter extraction and comparison of simulation and measurement at elevated temperatures is sparse.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00113-9

## 2. Experimental

We have investigated the MOS snap-back breakdown and its temperature dependence up to 300°C using LDD-NMOS transistors with effective gate lengths from 0.56μm up to 1.06μm and gate width of 20μm. The effective gate length is 0.14μm smaller than the nominal gate length. Fig. 1 shows the net doping profile on a horizontal cutline 0.1μm beneath the gate oxide. The investigations refer to a commercial twin tub epi-CMOS process. The devices were tested at wafer level on a hot chuck in a shielded probe station. Quasi static measurements were performed using the semiconductor parameter analyser HP 4155. The substrate current and drain voltage were recorded in dependence of the drain current up to breakdown with gate voltage as parameter for the temperatures 30°, 125°C, 250°C and 300°C.

## 3. Parameter extraction

Using the measured drain and substrate current we have extracted parameters for a simple lumped element model for simulation of snap back breakdown on circuit level. The lumped element model consisting of the NMOS transistor, a bipolar transistor, a current source for the avalanche multiplication and the substrate resistance is shown in Fig.2.

This model can be implemented easily into CAD tools for circuit design. Avalanche breakdown is modeled by the current and voltage dependent current source with the multiplication factor M which is given by

$$M = \left(1 - \exp\left\{K\left[\frac{1}{V_{AV}} - \frac{1}{V_{DS} - V_{dsat}}\right]\right\}\right)^{-1} \quad (1)$$

after [17,18], where $V_{AV}$ is the avalanche breakdown voltage, $V_{dsat}$ the saturation voltage of the MOS transistor, $V_{DS}$ the drain source voltage and K the exponential slope of avalanche breakdown. Using this approach, the dependence of the electric field at the drain junction on the gate voltage $V_{GS}$ is taken into account via its influence on $V_{dsat}$. $V_{dsat}$ was calculated from

$$V_{dsat} = \frac{v_{sat} L_{eff}}{\mu_s}\left[\sqrt{1 + \frac{2\mu_s (V_{GS} - V_{th})}{v_{sat} L_{eff}}} - 1\right] \quad (2)$$

after [19] with $v_{sat}$ saturation velocity, $\mu_s$ surface mobility in the channel and $V_{th}$ threshold voltage.

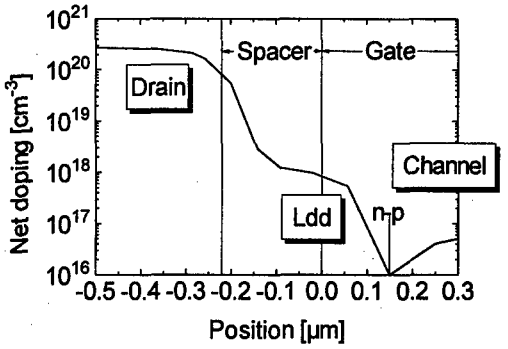



Fig. 1. Net doping profile on a horizontal cutline 0.1μm beneath the gate oxide.

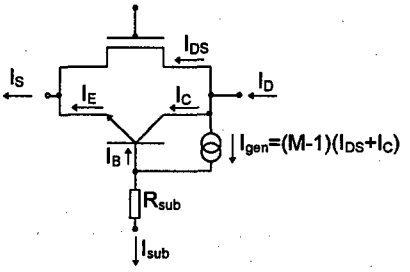



Fig. 2. Circuit model for simulation.

This equation is similar to, but slightly simpler than the equation used in SPICE Level 3 models [20]. $\mu_s$ and $V_{th}$ and there temperature dependence were given in the process document. $v_{sat}$ was calculated from

$$v_{sat} = \frac{2.4 \cdot 10^7 \, cm/s}{1 + 0.8 \exp(T/600K)} \quad (3)$$

after [21]. The calculation effort for $V_{dsat}$ can be avoided when using the curves for $V_{GS}$=0 for the extraction, although this is sometimes more dangerous due to possible breakdown of the gate oxide. Before the bipolar transistor is turned on, M is given by

$$M = \frac{I_D}{I_D - I_{sub}} \quad \text{and} \quad (4)$$

$$1 - \frac{1}{M} = \frac{I_{sub}}{I_D} = \exp\left\{K\left[\frac{1}{V_{AV}} - \frac{1}{V_{DS} - V_{dsat}}\right]\right\} \quad (5)$$

By plotting 1–1/M in logarithmic scale versus $1/(V_{DS}-V_{dsat})$, $V_{AV}$ and K can be extracted from a linear fitting curve as shown in Fig. 3. Using this model for the avalanche effect, the collector current $I_C$, its saturation current $I_S$, the bipolar gain $B_{npn}$ and the substrate resistance $R_{sub}$ can be extracted from that region of the IV curve, where the bipolar transistor is turned on.

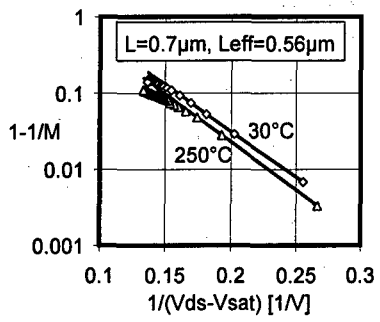



Fig. 3. Extraction of avalanche breakdown voltage and the exponential slope of avalanche multiplication by fitting.

$$I_C = \frac{I_D}{M} - I_{DS} \tag{6}$$

$$B_{npn} = \frac{I_D / M - I_{DS}}{(M-1) I_D / M - I_{sub}} \tag{7}$$

where $I_{DS}$ is obtained by extrapolating the drain current from the saturation region of the IV curve before avalanche multiplication occurs.

The parasitic bipolar transistor is modeled by equations used in PSPICE [23].

$$I_C = I_S \exp\left(\frac{qV_{BE}}{kT}\right) - \left(1 + \frac{1}{B_R}\right) \cdot I_S \exp\left(\frac{qV_{BC}}{kT}\right) - I_{SC} \exp\left(\frac{qV_{BC}}{kT}\right) \tag{8}$$

$$I_B = \frac{I_S}{B_F} \exp\left(\frac{qV_{BE}}{kT}\right) + I_{SE} \exp\left(\frac{qV_{BE}}{kT}\right) + \frac{I_S}{B_R} \exp\left(\frac{qV_{BC}}{kT}\right) + I_{SC} \exp\left(\frac{qV_{BC}}{kT}\right) \tag{9}$$

where $I_S$ is the transport saturation current, $I_{SE}$ the non-ideal base-emitter saturation current, and $I_{SC}$ the non-ideal base-collector saturation current. Due to the symmetry of the transistors we assume $B_F$=$B_R$ and $I_{SE}$=$I_{SC}$. The ideal maximum forward current gain $B_F$ can be calculated from the extracted current gain $B_{npn}$ from

$$B_F = \left(\frac{1}{B_{npn}} - \frac{I_{SE}}{I_S}\right)^{-1} \tag{10}$$

The temperature dependence implemented in the above equations is given by

$$B_F(T) = B_F \left[\frac{T}{T_{nom}}\right]^{XTB} \tag{11}$$

$$I_S(T) = I_S \left[\frac{T}{T_{nom}}\right]^{XTI} \exp\left(\left[\frac{T}{T_{nom}} - 1\right] \frac{qE_G}{kT}\right) \tag{12}$$

$$I_{SE}(T) = I_{SE} \left[\frac{T}{T_{nom}}\right]^{XTI - XTB} \exp\left(\left[\frac{T}{T_{nom}} - 1\right] \frac{qE_G}{kT}\right) \tag{13}$$

where XTI is the temperature exponent of the transport saturation current and XTB is the temperature exponent of the ideal maximum forward current gain BF and equivalent reverse gain BR. Unfortunately EG is assumed temperature independent in the equations given above in this version of PSPICE. This can be corrected by adding a value of 2.93 to XTI. The saturation current of a one sided abrupt $n^+p$ junction is given by

$$I_S = A \cdot \sqrt{\frac{k \cdot T \cdot \mu_{nB}}{q \cdot \tau_{nB}}} \cdot \frac{n_i^2}{N_A} \tag{14}$$

q is the electron charge, $\mu_{nB}$ the electron mobility with $\mu_{nB} \sim T^{BEX}$, BEX is between -1.5 and -2.42 depending on the scattering mechanism (after Sze [21] and $\tau_{nB}$ the electron lifetime in the base. Schmid and Reiner [26] found $\tau_n \sim T^2$. However, from the later described extraction of XTB we found $\tau_n \sim T^{0.9..1.82}$. The temperature dependence of the intrinsic density is given by $n_i \sim T^{1.5} \exp(-E_G/2kT)$. As a result the temperature dependence of the saturation current is then approximately given by $I_S \sim T^{1.4...2.3} \exp(-E_G/kT)$. With the correction mentioned above (+2.93) an XTI=4.3...5.3 was used to match the measured characteristics (XTI=5 fits best). The transport saturation current of the bipolar transistor can be extracted from

$$I_S = I_C \cdot \exp\left(-\frac{qV_J}{kT}\right) \text{ with } V_J\text{: junction volt.} \tag{15}$$

$$V_J = 0.7V \frac{T}{T_{nom}} + \frac{kT}{q} \ln\left[\frac{n_i^2(T_{nom})}{n_i^2(T)}\right] \tag{16}$$

at that point, where the bipolar transistor is turned on. $I_{SE}$=$I_{SC}$ can be estimated from the ratio of the doping concentrations in the p-well and LDD regions.

$$I_{SE} \approx I_S \frac{\mu_{p,LDD} \cdot N_{A,p-tub} \cdot L_{eff}}{\mu_{n,p-tub} \cdot N_{D,LDD} \cdot L_{LDD}} \tag{17}$$

$I_{SE}/I_S \approx 0.02$ at 30°C for $L_{eff}$=0.56µm. The substrate resistance can be obtained from:

$$R_{sub} = \frac{V_J}{I_{sub}} = \frac{kT}{qI_{sub}} \ln\frac{I_C}{I_S} \tag{18}$$

As was found by Scotnicki et al. [22] $R_{sub}$ is constant only up to a certain value of substrate current and then decreases rapidly with increasing $I_{sub}$ until it reaches an almost constant lower value. Fig. 4 shows the extracted substrate resistance versus

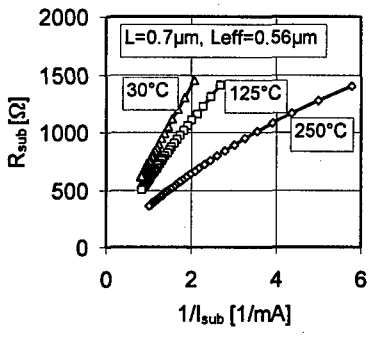



Fig. 4. Extracted substrate resistance versus reciprocal substrate current for $L_{eff}$=0.56µm at three different temperatures. Gate width is 20µm.

reciprocal substrate current for $L_{eff}$=0.56µm at three different temperatures. At high substrate current the substrate resistance is almost proportional to the reciprocal substrate current. To match the measured characteristics we used an average value of $R_{sub}$=1kΩ for simulations. Of course it would be more advanced to implement a current dependent resistance which will be the concern of further study. Scotnicki et al. [22] also found that the substrate resistance after bipolar turn on is almost independent of gate length and gate voltage.

All other parameters of the NMOS transistor were specified in the process document for the model BSIM3, version 2. The drain substrate leakage current is set to zero, as it is already accounted for in the bipolar model. In this work the extraction was done for $V_{GS}$=2V, as for measurements at $V_{GS}$=0 often a destruction of the transistors due to gate oxide breakdown occurred.

## 4. Circuit level simulation

Simulations were performed using MicroSim PSPICE A/D Version 6.3 [23]. To implement the dependence of M on $V_{GS}$ the PSPICE input file contains parameter definitions for the calculation of $V_{dsat}$ as a function of $V_{GS}$:

```
** Calculation of Vdsat (here named Vs)
.PARAM Vx={0.72-1.37e-3*(TEMP-30)}
.PARAM TR={(TEMP+273)/300}
.PARAM c={2.9*PWR(TR,1.8)/(1+0.8*EXP(TR/2))}
.PARAM Vs={c*(SQRT(1+2*(Vg-Vx)/c)-1)}
```

The most critical device for the simulations is the dependent current source describing avalanche multiplication. Due to its exponential dependence on drain voltage, numerical overflow can occur during calculations. we have prevented overflow by limiting M to the value of $10^{99}$ using logical expressions available within PSPICE:

```
**Exponential slope of avalanche multiplication factor
.PARAM K={28.05+(TEMP-30)/230}
.PARAM VAV={12.95+(TEMP-30)*6.14e-3}
** Simplification for calculation of M
.FUNC E(x)={IF(x>0,EXP(K*(1/VAV-1/x)),0)}
** Netlist
GAM 1 2 VALUE {IF(E(V(1)-Vs)<1-1e-99,
                  (I(VM)+I(VQ))*(1/(1-E(V(1)-Vs))-1),
                  1e99*(I(VM)+I(VQ)))}
```

## 5. Results and discussion

The measured IV-curves of an NMOS transistor with $L_{eff}$=0.56µm for 30°C, 125°C, 250°C and 300°C are shown in Fig 5 as solid lines. Fig. 6 shows the snap back sustaining voltage estimated from the measured IV-curves versus temperature with gate length as parameter. The snap-back sustaining voltage for $L_{eff}$=0.56µm increases from 8.25V at room temperature to 8.9V at 300°C. Table 1 gives the extracted parameters for the simulations for different temperatures and gate lengths. For 300°C the extraction failed due to the early turn on of the bipolar transistor. In this case we used linearly extrapolated values for the simulation.

Table 1
Extracted parameters for $V_{GS}$=2V

| $L_{eff}$=0.56µm | 30°C | 125°C | 250°C |
|---|---|---|---|
| $V_{dsat}$ [V] | 0.94 | 1.1 | 1.28 |
| $V_{AV}$ [V] | 12.95 | 13.5 | 14.3 |
| K [V] | 28.05 | 28.4 | 29 |
| $B_{npn}$ | 10 | 11.3 | 12.4 |
| $L_{eff}$=0.86µm | 30°C | 125°C | 250°C |
| $V_{dsat}$ [V] | 0.98 | 1.14 | 1.32 |
| $V_{AV}$ [V] | 13.42 | 14.21 | 15.39 |
| K [V] | 27.82 | 28.35 | 28.8 |
| $B_{npn}$ | 6.4 | 7.7 | 8.5 |

The avalanche breakdown voltage $V_{AV}$ for $L_{eff}$=0.56µm increases from 12.95V at 30°C to 14.3V at 250°C. This corresponds to a temperature coefficient of the avalanche breakdown voltage of ≈6mV/K. The exponential slope of avalanche multiplication K also increases with rising temperature. $V_{AV}$ is also dependent on gate length, probably due to different electric field distribution at the drain junction. The saturation currents were determined to $I_S=10^{-15}$A and $I_{SE}=2\cdot10^{-17}$A at 30°C.

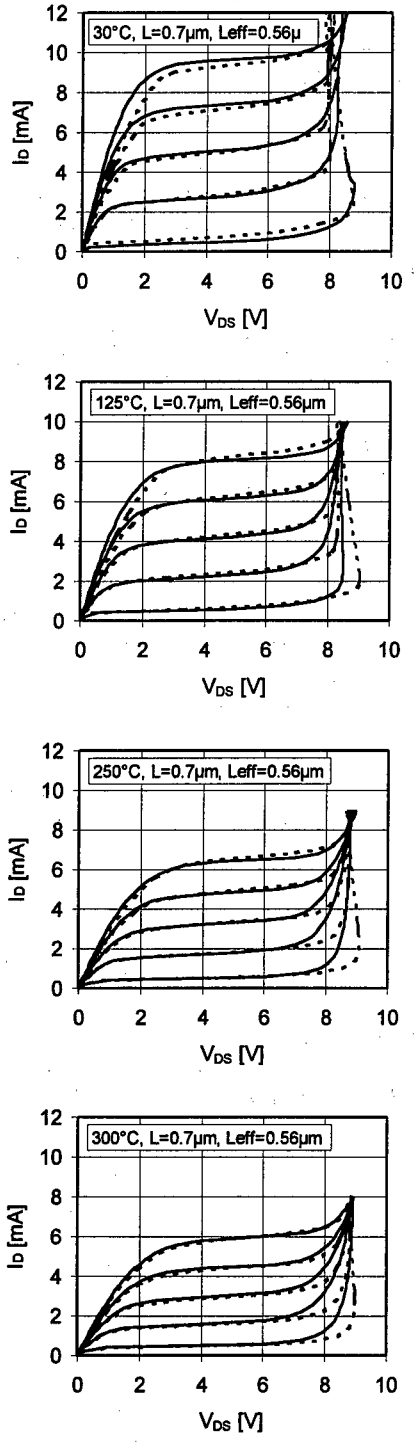



Fig. 5. Measured (solid lines) and simulated (broken lines) IV-curves of an NMOS transistor with $L_{eff}$=0.56µm for 30°C, 125°C, 250°C, and 300°C. Gate voltage varies in steps of 1V from 5V to 1V. Gate width is 20µm.

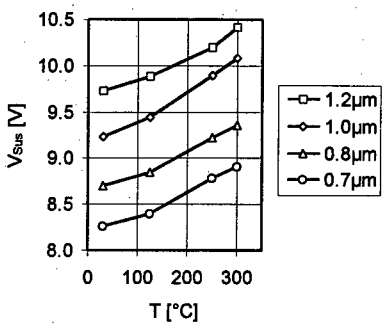



Fig. 6. Estimated sustaining voltage from measured IV-curves of snap back breakdown vs. temperature with nominal gate length as parameter.

The parasitic bipolar transistor is typically base transport limited, especially if $L_{eff}>0.1 \cdot L_{nB}$. This implies a much lower temperature dependence of current gain than for the injection limited case [23,24].

$$B_{npn} \sim \frac{1}{\cosh\left(\frac{L_{eff}}{L_{nB}}\right)-1} \approx \frac{2L_{nB}^2}{L_{eff}^2} = \frac{2kT\mu_{nB}\cdot\tau_{nB}}{qL_{eff}^2} \quad (19)$$

after [20] with $\mu_{nB} \sim T^{BEX}$, BEX=-1.5...-2.42 depending on the scattering mechanism [21] and $\tau_{nB} \sim T^2$ [26]. Then the temperature dependence of $B_{npn}$ is given by $B_{npn} \sim T^{0.58...1.5}$, which is higher than the experimentally extracted temperature exponent of approximately 0.4. Therefore, the temperature exponent of $\tau_n$ is assumed to be smaller than 2, probably between 0.9 and 1.82. An analysis of the gate length dependence of $B_{npn}$ shows that $B_{npn}$ is rather proportonial to $1/L_{eff}$ than to $1/L^2_{eff}$. Probably high injection effects result in a slower increase of $B_{npn}$ with decreasing gate length than expected.

An analytical expression for the sustaining voltage $V_{sus}$ can be derived from the region where the bipolar transistor is turned on and the substrate current is negligible compared to the base current. This corresponds to an open base mode, where $(M-1)\cdot B_{npn}=1$ is the condition for sustaining snap back, giving

$$V_{sus} = \frac{1}{\frac{1}{V_{AV}} + \frac{1}{K}\ln(1+B_{npn})} + V_{dsat} \quad (20)$$

Using this equation with the extracted parameters results in values for $V_{sus}$ more than 1V smaller than the experimentally observed values. This difference is known from literature [17] and has been treated by substituting $B_{npn}$ by $B_{npn}/k$, where k=1..2 is a fitting parameter which is determined to match the

experimental snap back voltage. The origin of k>1 results from the fact that the influence of the substrate current and substrate resistance on drain voltage may not be neglected. An open base mode is not achieved because the substrate resistance becomes smaller with increasing substrate current (rather grounded base than open base). Nevertheless, the calculated values for $V_{sus}$ show the same temperature dependence as the experimentally observed values. Therefore, Eq. 20 is useful for qualitative interpretation and estimations.

The simulated and measured IV-curves match quite acceptable (see Fig. 5). Deviations in the breakdown region are probably due to neglecting the current dependence of the substrate resistance. Just at the onset of the bipolar transistor the substrate resistance is higher than assumed in the simulations leading to the observed lower measured drain voltage in this section of the breakdown curve.

## 6. Conclusion

We conclude that the increase of avalanche breakdown voltage and exponential slope of the avalanche multiplication factor compensate the increase in bipolar gain leading to an increase of snap back sustaining voltage with temperature. If the extracted parameters describing snap back would be specified in process documents, circuit designers could use them to identify and solve problems related to both ESD protection circuits and EOS. The results are also relevant for high temperature operation of electronics, which is a performance issue of growing importance. Further work implementing a current dependent substrate resistance should result in a better matching of simulated and measured curves and accurate calculation of substrate current.

## References


[1] J. Stemmer, J. Ackermann, D. Uffmann, and J. Aderhold, *Proc. 3rd Int. High Temp. Electron. Conf.*, Albuquerque, New Mexico, USA, VIII-9 (1996).

[2] J. Ackermann, J. Stemmer, D. Uffmann, and J. Aderhold, *Proc. 3rd Int. High Temp. Electron. Conf.*, Albuquerque, New Mexico, USA, P-59 (1996).

[3] D. Uffmann, J. Stemmer, H.-U. Schröder, J. Ackermann, and J. Aderhold, Proc. **SPIE 2636**, 32 (1995).

[4] P.L. Dreike, D.M. Fleetwood, D.B. King, D.C. Sprauer, and T.E. Zipperian, *IEEE Trans. on Comp., Packaging and Manufact. Technol.- Part A*, **17**, 4, 594 (1994).

[5] T. Polgreen and A. Chatterjee, *EOS/ESD Symp. Proc.*, 167 (1989).

[6] R.R. Troutman, *IEEE Trans. on Electron Dev.*, **ED-23**, 4, 419 (1976).

[7] T. Toyabe, K. Yamaguchi, S. Asai, and M.S. Mock, *IEEE Trans. on Electron Dev.* **ED-25**, 7, 825 (1978).

[8] E. Sun, J. Moll, J. Berger, and B. Alders, *IEDM Tech. Dig.*, 478 (1978).

[9] N. Khurana, T.J. Maloney, and W. Yeh, *IEEE IRPS*, 212 (1985).

[10] A. Amerasekera and C. Duvvury, *ESD in silicon integrated circuits*, p. 115ff. John Wiley & Sons, Chichester (1995).

[11] C. Duvvury and A. Amerasekera, *Proc. of the IEEE*, **81**, 5, 690 (1993).

[12] G.J. Gaston, B.S. Bold, and J.B. Mason, *Proc. IEEE Int. Conference on Microelectronic Test Structures*, 7, 28 (1994).

[13] M. Reisch, IEEE Trans. on Electron Dev. **39**, 6, 1398- 1409 (1992).

[14] G. Verzellesi, G. Baccarani, C. Canali, P.Pavan, L. Vendrame, E. Zanoni, IEEE Trans. on Electron Dev. **40**, 12, 2296-2300 (1993).

[15] C. Diaz and S.-M. Kang, Trans. on CAD, **11**, 11, 1344-1354 (1992).

[16] C. Diaz, S.-M. Kang, C. Duvvury, IEEE Trans. on CAD of IC and Systems, **13**, 4, 482-493 (1994).

[17] A. Amerasekera, S. Ramaswamy, M.-C Chang, C. Duvvury, 34th Int. Rel. Phys. Symp. Proc., 318-326 (1996).

[18] A. Amerasekera, M.-C Chang, C. Duvvury, *Circuits and Devices*, Simulation & Modeling, 7-10 (3/1997).

[19] R.S.U. Dike, *Invest. of the temp. depend. of MOSFET mechanisms and charact. and their consequ. on CAD of VLSI circ. design and operat.* Ph.D.Thesis, Univ. of Birmingham, UK, (1986).

[20] P. Antognetti, G. Massobrio, *Semiconductor device modeling with SPICE*, McGraw-Hill, New York (1988), p.190.

[21] S.M. Sze, *Physics of semiconductor devices*, 2nd Edn, John Wiley and Sons, New York (1981).

[22] T. Scotnicki, G. Merckel, C. Denat, Solid-State Electronics **35**, 5, 717-721 (1992).

[23] MicroSim PSPICE A/D Reference Manual, MicroSim Corporation, 20 Fairbanks, Irvine, CA 92718 (1996).

[24] Kauffman and A.A. Bergh, IEEE Trans. on Electron Dev. ED-15, 10 (1968) 732-735.

[25] Martinelli, IEEE Trans. on Electron Dev. ED-23, 11 (1976) 1218-1224.

[26] W. Schmid and J. Reiner, *J. Appl. Phys.* **53**, 9, 6250-6252 (1982).

PERGAMON

Microelectronics Reliability 38 (1998) 1139–1141

MICROELECTRONICS RELIABILITY

# Modelling the Field Soft Error Rate of DRAMs by varying the critical cell charge

Horst Schleifer [a], Oskar Kowarik [a], Kurt Hoffmann [a], Werner Reczek [b]

[a] *University of Bundeswehr Munich, Neubiberg, Germany*
[b] *Siemens, Munich, Germany*

**Abstract**

The Field Soft Error Rate (FSER) has been determined by the variation of the critical charge and the measurement of the charge collection volume determined by alpha-particle irradiation. The modelled FSER versus critical charge dependence agrees well to the one of the Field Soft Error measurements. The results futher show, that the impact of the on-chip Alpha-Particle flux can be neglected.

© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Determining the Field Soft Error Rate of DRAMs is a major problem for manufacturers of memory products. Reliable data can only be gained in the field by the evaluation of large numbers of memory chips. This problem can be avoided by using an Accelerated Field Soft Error Rate test (AFSER) (Fig. 1). The flux of secondary particles of the cosmic radiation is currently modelled by high energy neutrons or protons (Fig. 1a) and the on-chip alpha-particles are modelled by use of an alpha-particle source (e.g. Thorium) (Fig. 1b).

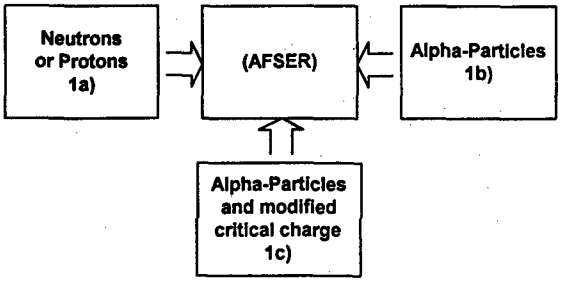



Fig. 1: Accelerated Field Soft Error Rate tests.

To our knowledge, the only reliable cosmic radiation modelling is performed by the use of neutrons or protons [1] at the disadvantage of high cost. Therefore, there is a need to come up with a new, inexpensive and simple method.

## 2. Modelling the impact of cosmic radiation by varying critical cell-charge

The proposed method (Fig. 1c) uses a single energy alpha-particle source and simultaneously a variable dummy-cell- and cell-charge. The variation of the dummy-cell charge is performed by varying the dummy-cell voltage $V_{DC}$ and the charge of the cells by injection of light (Fig. 2). The measurement setup is shown in Fig. 3.

The purpose is to generate a defined critical charge $Q_c$. This charge $Q_c$ is used to determine the charge-sensitive area of the memory chip by using an external alpha-source. The charge-sensitive area $\sigma$ is defined by the ratio of the measured Accelerated Soft Error Rate (ASER) as a function of the critical charge $Q_c$ with respect to the alpha-flux $I_\alpha$ of the alpha-particle source.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00108-5

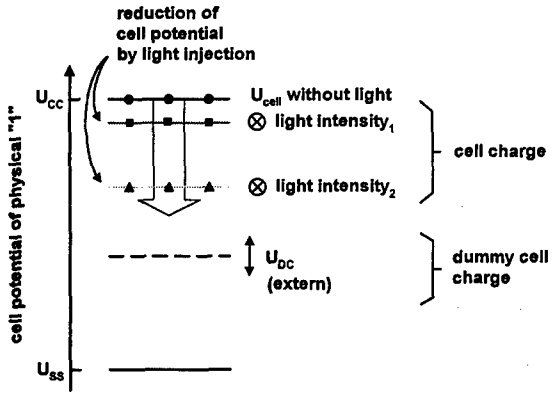



Fig. 2: Variation of the critical charge $Q_c$.

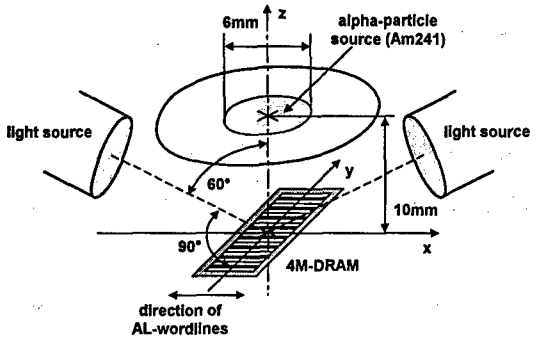



Fig. 3: Measurement setup.

$$\sigma = \frac{\mathrm{ASER}(Q_c)}{I_\alpha} \tag{1}$$

This sensitive area increases if $Q_c$ is reduced (Fig. 4) and can be described by a Weibull-distribution.

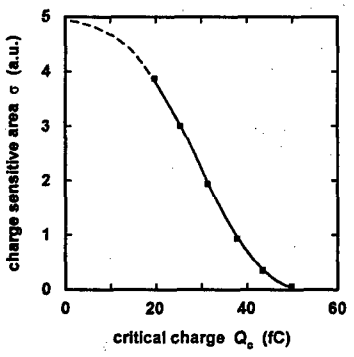



Fig. 4: Charge sensitive area.

Since $Q_c$ is known, the minimum charge collection depth $d_{col}^{\min}$ for the respective sensitive area can be determined by

$$d_{col}^{min} = \frac{Q_c}{LQT_\alpha} \tag{2}$$

which includes the linear charge transfer of an alpha-particle $LQT_\alpha \approx 60 \cdot 10^3$ charge carriers per micron. This results in an effective collection volume of

$$V_{col} = \int_0^\infty d_{col}^{min} \cdot \left( \frac{\partial \sigma(d_{col}^{min})}{\partial d_{col}^{min}} \right) \cdot \partial d_{col}^{min} = \int_0^\infty \sigma(d_{col}^{min}) \cdot \partial d_{col}^{min}$$

$$V_{col} = \frac{1}{LQT_\alpha} \cdot \int_0^\infty \sigma(Q_c) \cdot \partial Q_c \tag{3}$$

Using this volume and the critical charge $Q_c$ under normal operation conditions the FSER can be calculated by

$$SER = V_{col} \cdot \mathrm{IBR}(Q_c) \tag{4}$$

where IBR($Q_c$) is the Integrated Burst Rate, depending on the critical charge $Q_c$ and can be found in Ziegler [2].

## 3. Modelling the impact of the alpha-particle flux

For evaluating the impact of the on-chip alpha-particles (Fig. 1b), that have there origin in trace amounts of radioactive isotopes (e.g. uran, thorium and other possible contaminations) as well as of bursts of atoms caused by energetic secondary particles of cosmic radiation, samples of each FSER test lot are analysed by alpha irradiation (ASER measurement). Fig. 5 ① shows the measurement results in dependence on the measured critical charge of the samples of the test lots, whereas Fig. 5 ② shows the dependence on the critical charge for a single device.

The contribution of the alpha-particles to the FSER is expected to be proportional to the measured ASER. Therefore a only alpha-particle caused FSER would be expected to change its value for the measured test lots over approximately 4 decades as shown on the right-side scale of Fig. 5.

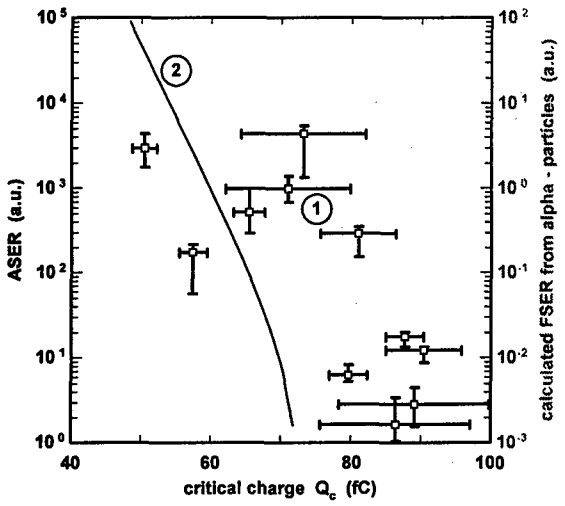



Fig. 5: ASER as function of critical charge.

## 4. FSER Measurement results

A field test has been performed with 11 lots of 3000 parts each for several weeks in order to determine the FSER. Samples of each lot were analysed by alpha-irradiation (Fig. 5). The results are compared in Fig. 6.

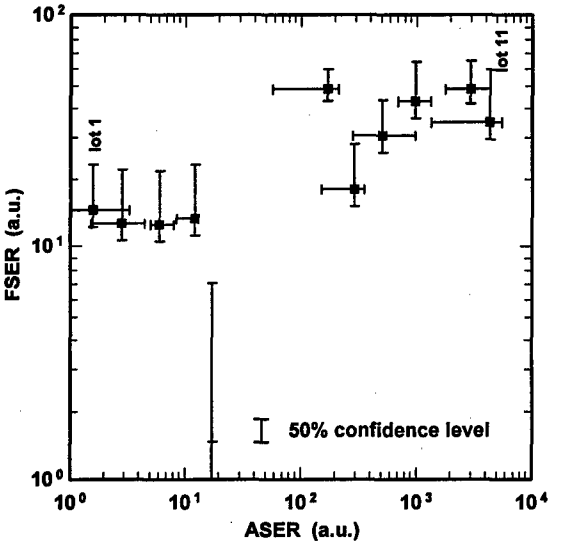



Fig. 6: Correlation between FSER and ASER.

The dots show the correlation between the measured FSER and the measured ASER. This figure shows furthermore a FSER spread of less than one decade whereas the ASER has a spread of approximately 4 decades. This means that the ASER measurement is not a useful model for predicting the FSER.

Therefore the FSER of the lots shown in Fig. 6 has been correlated to the critical charge $Q_c$ (Fig. 7) and is shown as black dots. This result allows the comparison of the modeled FSER (Eq. 4) using the data of the critical charge.

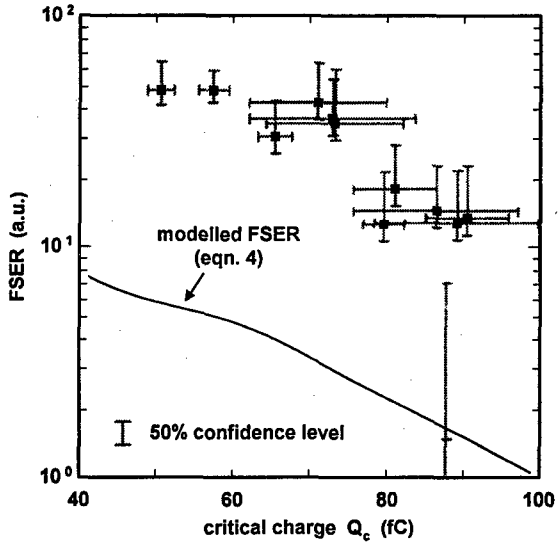



Fig. 7: FSER as a function of critical charge.

The discrepancy between measurements and theory (Eq. 4) is due to the definition of the burst rate covering only bursts of less than about 200nm in diameter.

A comparison of the FSER results in Fig. 7 with the ASER results in Fig. 5 shows no impact of the alpha-particles on the FSER.

In summary it can be said that the theory derived allows the prediction of the FSER by simply measuring the critical charge and the charge collection volume and by using the point burst rate of [2].

## References

[1] J. F. Ziegler, H. P. Muhlfeld, C. J. Montrose, H. W. Curtis, T. J. O'Gorman, J. M. Ross "Accelerated testing for cosmic soft-error rate," IBM J. Res. Develop., vol. 40, no. 1, 1996.

[2] J. F. Ziegler, W. A. Lanford, "The effect of sea level cosmic rays on electronic devices," J. Appl. Phys., 52(6), pp. 4305-4312, June 1981.



PERGAMON

Microelectronics Reliability 38 (1998) 1143–1148

MICROELECTRONICS RELIABILITY

# Design of a Low EMI Susceptibility CMOS Transimpedance Operational Amplifier

Gianluca Setti, Nicolò Speciale

D.E.I.S. - Università di Bologna, viale Risorgimento 2, 40136 Bologna - ITALY

e-mail: (gsetti,nspeciale)@deis.unibo.it

**Abstract**

In this work we analyze the effects of electromagnetic–induced interferences conveyed at the input of a transimpedance CMOS operational amplifier. In particular, it will be highlighted that transimpedance amplifiers natural exhibit a lower EMI susceptibility compared to common voltage-feedback opamps. Moreover, it will be shown through simulations that a careful circuit design can lead to opamps with a practically vanishing EMI susceptibility.

© 1998 Elsevier Science Ltd. All rights reserved.

## 1 Introduction

Since the first experiments in radiocommunication, it has been observed that spark gaps generate electromagnetic waves which are very rich in spectral components and which are therefore likely to cause interference in electronic systems. Nowadays, several other sources of electromagnetic emission exist (ranging from radio transmitters and radars to relays and dc electric motors, as well as digital electronic devices), so that the problem of designing electromagnetic compatible (EMC) systems has become of great practical concern. A basic situation of interest for electromagnetic interference (EMI) analysis can be schematically represented as an electromagnetic radiation emitted by a given source, which is transferred though a coupling path to an electronic device, where it is processed, potentially result-

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00109-7

ing in undesired behavior. For instance, EM radiation can incide upon the outer enclosure of an apparatus and be coupled through skin apertures to its interior. The resulting internal EM fields induce RF voltages on the system cables which are conducted to the terminals of circuits and semiconductor devices located inside the electronic equipment, which thus act as unintentional receivers.

EMI effects were first analyzed with reference to military applications, especially in avionic and aerospace fields, but currently, due to the large adoption of electronic and microelectronic equipments, EMI are experimentally evaluated and carefully studied to find possible prevention methodologies for practically all electronic systems [1-8]. Power-supply filters, shielded cables and shielded and filtered connectors are common, though expensive, solutions to increase system EMC.

Nowadays, to increase reliability and decrease system cost and size, a great effort is also given to the design of ICs containing on the same chip both the digital and the analog part. This has a twofold consequence on the EMC design of these circuits. On one hand, the interference effects between embedded systems tend to become more important, due for instance to the presence of the integrated high–frequency oscillator needed for generating the clock signal for the digital part of the IC. On the other hand, the above mentioned solutions to increase system EMC cannot be employed, so that it appears necessary to develop some guidelines in the *design phase* to reasonably assure a sufficient amount of system robustness against EMI.

Additionally, as CMOS is the leading technology for most digital applications, a great interest has been and must be devoted to the design of high–performance mixed analog/digital CMOS ICs, which may contain several operational amplifiers (opamps).

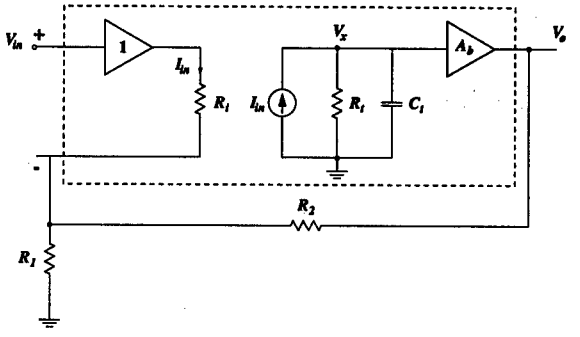



Figure 1: Model of a transimpedance amplifier connected in a non-inverting configuration

Among the many proposed opamps structures, voltage feedback opamps (VOAs) are extensively employed in electronic applications as versatile analog building blocks. Although many different topologies, in CMOS, bipolar or BiCMOS technology, have been proposed in the last twenty years to increase the circuit performances, VOAs have the intrinsic limitation of being characterized by a constant gain-bandwidth product. Moreover, a further drawback of this circuit topology is the difficulty in obtaining an high slew-rate value.

Conversely, current feedback opamps (CFOAs) do not suffer from this limiting factors [9]. Figure 1 shows a simplified model of a transimpedance amplifier connected in a closed-loop non-inverting configuration. As can be deduced from the scheme, in a CFOA the inverting input is at low impedance value ($R_i$), so that current feedback can be applied, while the non-inverting input is connected to a voltage buffer to achieve an high input impedance value. The input current $I_{in}$ is then transferred trough a current mirror to an high impedance node represented by resistance $R_t$ and capacitance $C_t$. This voltage is then transferred to the output by a voltage buffer $A_b$, providing a low value for

the opamp output impedance. By straightforward computations, the voltage gain for the amplifier of Figure 1 can be expressed as

$$\frac{V_o}{V_{in}} = \frac{G}{1 + \frac{R_2 + R_i G}{A_b R_t} + j\omega C_t \frac{R_2 + R_i G}{A_b}} \quad (1)$$

where $G = 1 + R_2/R_1$ is the ideal closed loop gain. The pole frequency is thus given by

$$f_p = A_b/(2\pi C_t(R_2 + R_i G)) \quad (2)$$

which, for $R_2 \ll R_i G$, directly depends only on $R_2$, so that the closed loop bandwidth is almost independent of the closed-loop gain. Moreover the slew-rate of a CFOA is extremely high. In fact a change $\delta V_{in}$ at the non-inverting input of the amplifier will be copied by the input buffer to the inverting input, thus changing the current through $R_1$ by $\delta V_{in}/R_1$ while the current through $R_2$ changes by $\delta V_{in}/R_2$, since $V_o$ initially remains constant. The total current change in the feedback resistors must be supplied by the input buffer and is expressed by $\delta I_{in} = \delta V_{in}(R_1 + R_2)/R_1 R_2$, which is copied to the high impedance node, thus causing a rapid change in $V_x$ and then in $V_o$. Therefore, in principle, the larger the input voltage slew-rate, the larger $\delta I_{in}$, the faster the change in the output voltage, while, in practice, the maximum slew-rate will be limited by the ability of the power supply to deliver a sufficient amount of current and of the non-ideal behavior of the current mirrors. Typical CFOAs exhibit a slew-rate value of several hundreds of V/$\mu$s. As shown in [5,6,7,8] the large-signal behavior, and in particular the slew-rate performance of an opamp, are deeply linked to the circuit EMI susceptibility to interference. Therefore, due to virtual no slew-rate limit of transimpedance amplifiers it is natural to expect that they will exhibit a significant EMI susceptibility at frequencies of the input signals much higher than those found in VOAs.

In this context, we analyze the failures induced from EMI conveyed–signals to the input of a CMOS CFOA architecture. In particular, we will show that susceptibility to EMI observed in this amplifier is associated with both asymmetries observed in opamp behavior during transients and to the value of some unavoidable parasitic capacitances in the input buffer stage. Finally, we will show the design–feasibility of a transimpedance integrated CMOS opamps with both good general performances and very low EMI susceptibility.

## 2 Design of a transimpedance CMOS opamp with low susceptibility to EMI

Figure 2 shows the CMOS transimpedance opamp topology analyzed in this work. Transistors $M_5$-$M_6$ and $M_9$-$M_{10}$ represent a class AB complementary common-drain input buffer, which provides an high impedance non-inverting input and copies $V_{in}(V^+)$ to the low impedance inverting input node. The drain current of $M_9$ and $M_{10}$ are transferred to the high impedance node 14 through a complementary current mirroring stage implemented trough transistors $M_8$ and $M_{11}$-$M_{19}$. Since the amplifier transimpedance is directly linked to the impedance at node 14, a regulated cascode current mirror topology has been employed, to assure a particularly high value. Finally, the drain voltage of $M_{14}$ and $M_{16}$ is then transferred to the output by the output voltage buffer $M_{21}$-$M_{22}$ and $M_{24}$-$M_{25}$.

As demonstrated in our previous works [4,5,6,7,8], devoted to the design and analysis of several bipolar, JFET/bipolar, CMOS

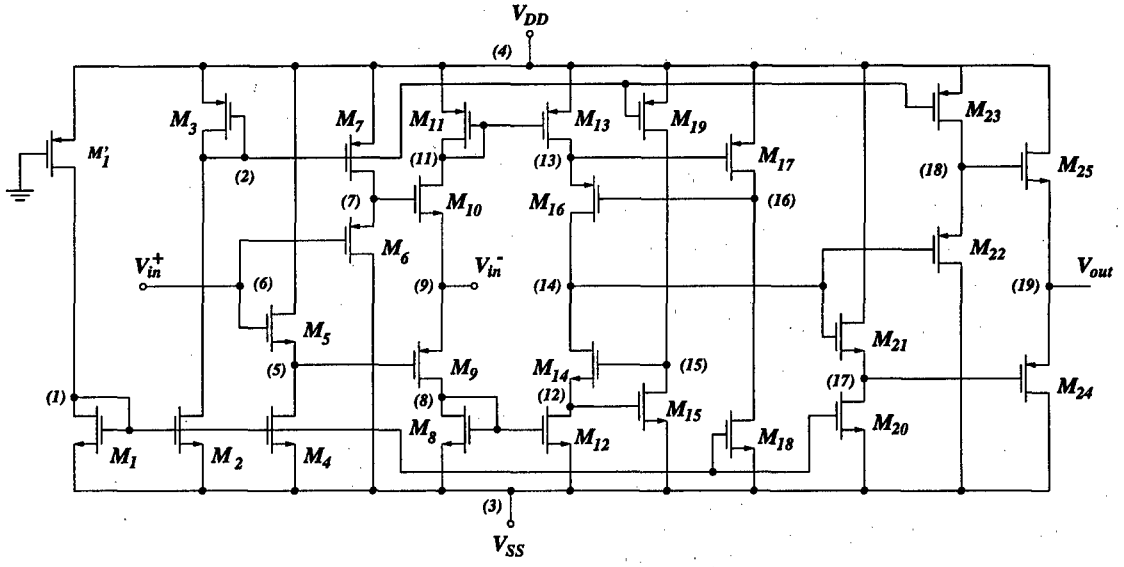



Figure 2: Complete schematic of the designed transimpedance CMOS amplifier

and BiCMOS VOAs, opamps can evidence a low EMI–susceptibility only if they have a very symmetric transient response. Thus, for the opamp of Figure 2 we devoted particular care to investigate the cause of the difference between positive and negative slew-rate value, which is mainly due to the asymmetry in the transient response in the two complementary common-drain stages of the input buffer. We found that such an asymmetry can be compensated by adding a suitable capacitance $C_b$ between nodes 9 and 14. By so doing and using a well–assessed $1.2\mu m$ CMOS process, we were able to design a 5 V CMOS transimpedance opamp with a dc transimpedance value $(V_o/I_{in})$ of 120 dB a gain-bandwidth of 300 MHz, a positive slew-rate of $462\,V/\mu s$, a negative slew-rate of $-452\,V/\mu s$, (with $C_b = 60fF$) and with a total power consumption at 11 mW.

In previously related papers of our group [6,7,8] it is shown, from both an experimental and a circuit analysis point of view, that among the several inverting and non–inverting amplifier configurations one can build with an opamp, the one which undergoes the strongest EMI influence is the non-inverting topology, particularly with respect to low values in the voltage gain $G$. Thus, in this work we will refer to a non-inverting configuration with $G = 2$ only, since similar results can be found for higher values of $G$.

The effects of an electromagnetic interference conveyed to the input terminals of the integrated opamp were simulated by applying at the input of the non-inverting amplifier a sinusoidal voltage of several amplitudes and frequencies. In particular, the results reported in the following figures refer to 400 mV and to frequencies up to a few GHz, i.e. up to values which are much higher than both the gain–bandwidth and the power–bandwidth PBW of the opamp. Of course, for an ideal opamp the application of this kind of signals should give output voltages with a zero DC value.

Figure 3 shows the simulated DC behaviour of the output voltage against frequency of the input signal for the opamp in Figure 2 for $C_b = 0, 60, 140$ fF. As can

be seen, the opamp evidences a significant susceptibility to EMI starting from frequencies of the order of the gain-bandwidth for $C_9 = 0$ (continuous-line). Conversely, when the topology with more symmetric transient response was considered ($C_b = 60, 140$ fF) we get the significantly better results (dashed- and dotted-line). Worthwhile to stress that EMI susceptibility becomes appreciable at much higher frequencies with respect to those found in VOAs [4-8]. However, even in such a case the opamp evidences a significant EMI susceptibility at very high frequencies, whose maximum value is always positive and increasing in magnitude with the magnitude of the input signal.

To explain these results, we carefully reanalyze the influence of parasitics on the input buffer. In particular we found that the capacitances at nodes 9 and 14 play critical role during transients, when a high-frequency out-of-band large-voltage interfering signal is applied to the opamp input.

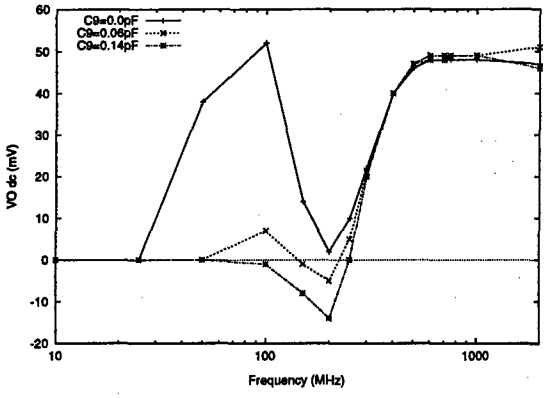



Figure 3: Mean value of $V_o$ against frequency of the interfering signal for $C_b = 0, 60, 140\, fF$

On the basis of the above analysis we were able to develop a second additional compensation network consisting of an additional capacitance $C_z$ connected between gate and source terminal of $M_5$ which permits to get an opamp with both satisfactory general characteristics and a low susceptibility to high frequency interfering signals. Some of the achieved results are reported in Figure 4 (a)-(c), which refers to $C_b = 0, 60, 140\, fF$ respectively and where $C_z = 325\, fF$. Figure 4 (a) clearly indicate that the role of $C_b$ is of decreasing EM susceptibility of the amplifier against "intermediate" frequency interfering signals. Moreover whenever $C_z$ is employed, from Figure 4 (b)-(c) we get practically zero-influence of the EMI signals up to several hundred of MHz. Similar results were achieved for other values of the gain of the opamp.

## 3 Conclusion

In this work the design of a transimpedance operational amplifier with low-probability EMI induced failure has been reported. In particular, we have found that the magnitude of the DC component at the output of the opamp due to sinusoidal input signals representing electromagnetic interference conveyed to the input terminals is correlated with the symmetry of the slew rate and parasitic capacitances. Additionally, we proposed a very simple compensation network to get an opamp structure with good performances and very low EMI susceptibility up to several hundred of MHz.

## References

[1] J. G. Tront, J. J. Whalen, C. E. Larson, J. M. Roe, "Computer–aided analysis of RFI effects in operational amplifiers", *IEEE Trans. Electromagnetic Compatibility*, Vol. EMC–21, pp. 297–306, 1979.

[2] J. J. Whalen, "Determining EMI in Microelectronics - A Review of the Past Decade",

*Proc. of 6th Symposium and Technical Exhibition on Electromagnetic Compatibility*, pp. 337–344, 1985.

[3] J. G. Tront, "Predicting URF upset in MOSFET digital IC's", *IEEE Trans. Electromagnetic Compatibility*, Vol. EMC–27, pp. 64–69, 1985.

[4] D. Golzio, S. Graffi, G. Masetti, "New circuit modeling of operational amplifiers", *IEEE Trans. on Electromagnetic Compatibility*, Vol. EMC–33, pp. 26–34, 1991.

[5] D. Golzio, S. Graffi, G. Masetti, "New macromodel for the analysis of EMI effects in LF355 opamp circuits", *Proc. of 1989 Int. Symposium on Electromagnetic Compatibility*, Nagoya, Japan, 1989.

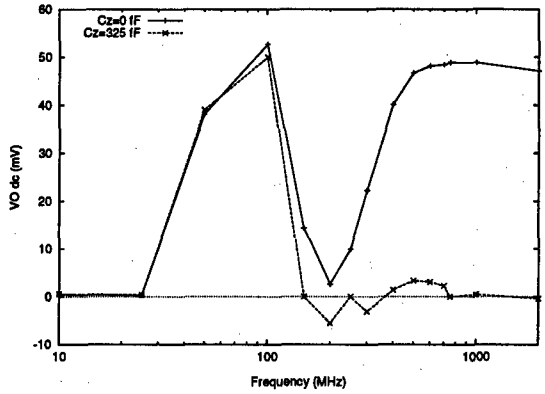
[6] D. Golzio, S. Graffi, Zs. M. V.–Kovàcs, G. Masetti, "Circuit macromodels and large–signal behaviour of FET–input operational amplifiers", *Int. J. of Circuit Theory and Applications*, Vol. 20, pp. 75–85, 1992.

[7] G. Masetti, S. Graffi, D. Golzio, Zs. M. V.-Kovàcs, "Failures induced on analog integrated circuits from conveyed electromagnetic interferences", Invited paper, *Proc. of 1993 ESREF Conference*, pp. 363–370, Bordeaux, 4–7 Oct. 1993.

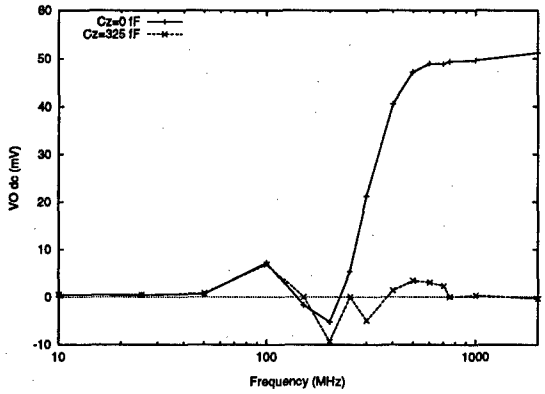
[8] P. Mattei, S. Graffi, Zs. M. V.-Kovàcs, G. Masetti, "Design of Integrated BiCMOS Operational Amplifiers with Low–Probability EMI–Induced Failures", *Microelectronics and Reliability*, Vol. 35, pp. 567–586, 1994.

[9] J. Huijsing, "Operational Floating Amplifiers", *Proc. IEE*, vol. 137, n. 2, part G, pp. 131-136, 1990.

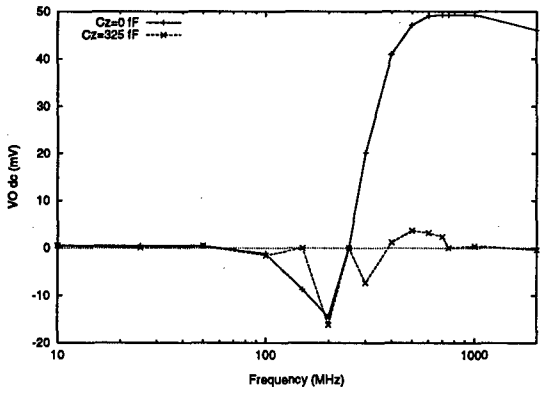
[10] R. Hogervorst, J. P. Eschauzier, J. H. Huising, "A Compact Power–Efficient 3V Rail–to–Rail Input–Output Operational Amplifier for VLSI Cell Libraries", *Proc. of Int. Solid-State Circuits Conference*, pp. 224–226, 1994.




(a)




(b)




(c)

Figure 4: Mean value of $V_o$ against frequency of the interfering signal for $C_z = 0$ and $C_z = 325\,fF$ with $C_b = 0\,fF$ (a), $C_b = 60\,fF$ (b) and $C_b = 140\,fF$ (c) respectively

PERGAMON

Microelectronics Reliability 38 (1998) 1149–1153

MICROELECTRONICS RELIABILITY

# Gate bias stress in hydrogenated and unhydrogenated polysilicon thin film transistors

B. Tala-Ighil[a], H. Toutah[a], A. Rahal[b,c], K. Mourgues[b], L. Pichon[b], F. Raoult[b], O. Bonnaud[b], T. Mohammed-Brahim[a,b]

[a] *Site Universitaire, LUSAC, BP78, 50130 Octeville, FRANCE,*

[b] *Groupe de Microélectronique et Visualisation, UPRESA au CNRS 6076, Université de RENNES I, Campus de beaulieu, 35042 Rennes cedex, FRANCE*

[c] *LCMS, Institut de Physique, USTHB, BP32 El-Alia, Alger, ALGERIE*

**Abstract**

Polysilicon Thin Film Transistors (TFT's), fabricated at temperature lower than 600°C, are now largely used in many applications, particularly in large area electronics. The reliability of these TFT's under different electrical conditions is then questionable. In this work, Gate bias stress is studied in two types of polysilicon TFT's originated from the same process. One type is unhydrogenated and the other is submitted to a Radio-Frequency hydrogen plasma. As this hydrogenation step is known to improve the TFT's performances but to introduce unstability, the unhydrogenated TFT's are expected to be more stable. The behaviours of the two types of TFT's under the gate bias stress are found however only different. The bias aging of unhydrogenated TFT's fit with the known model of the n-channel c-Si MOSFET's bias stress. The behaviour of the hydrogenated TFT's is explained from the model of defect creation in hydrogenated amorphous silicon.

© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Low temperature polysilicon Thin Film Transistors (TFTs) are of a great interest for many applications of polysilicon technology such as CMOS, RAM technologies and more particularly for large-area electronics. However, Polysilicon TFT's are post-hydrogenated to passivate the in-grain boundaries and in-grain defects. As the unstability of a-Si:H films originates principally from the behaviour of hydrogen in the material, numerous studies of the stability of hydrogenated polysilicon TFT's were done in the last few years. With the recent advent of high performance unhydrogenated polysilicon TFT's realized in low temperature Solid-Phase Crystallization Technology in the GMV (field effect mobility higher than 100 $cm^2V^{-1}s^{-1}$, subthreshold slope around 0.6 V/dec) [1], their stability is expected to be improved. In this study, the effect of gate bias stress on these SPC TFT's before and after RF plasma hydrogenation is studied.

## 2. Devices and stress conditions

Silicon films, 300 nm thick, are deposited in a conventional horizontal hot-wall Low Pressure Chemical Vapor Deposition (LPCVD) reactor using pure silane as source gas. The glass substrate, 2x2 square inch, is covered with 200 nm deposited APCVD silicon dioxide. The first 150 nm of the 300 nm silicon film are undoped by using only pure silane gas. The last 150 nm are n-type doped by switching on the phosphine opening valve during the deposition. This structure is crystallized and then processed to fabricate n-type TFTs, in the configuration presented in Figure 1. The doped up-layer is plasma etched to define channel, source and drain regions. Therefore, a 50 nm thick $SiO_2$ is

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00098-5

deposited at 450°C by APCVD technique to ensure the gate insulation. It is then thermally annealed in $N_2$ at 600°C to ensure a good densification and to eliminate any water related mobile ion effect. Source and drain contacts are opened through the gate insulator. Finally, aluminum is thermally evaporated and wet etched to form source, drain and gate electrodes. Post-metallization annealing is performed at 390°C in an atmosphere of forming gas

The glass wafer is then cut in two parts. One part is 13.56 MHz plasma hydrogenated at 300°C. The duration of the hydrogenation process depends on the transistor dimension particularly on the channel length. The dependence was studied in previous works [2] and is due to the lateral diffusion time of atomic hydrogen along the channel. The hydrogenation duration is then choosen to 3 hours for the present TFT's where the channel lenght and width values are 22 μm and 60 μm respectively.

TFT's are characterized at 25°C and then 20V positive or negative gate bias is applied at the same temperature. The 20V bias is applied to avoid the Fowler-Nordheim injection of carriers in the gate oxide. The source and drain are short circuited during the gate bias stress to avoid any hot-carrier effect.

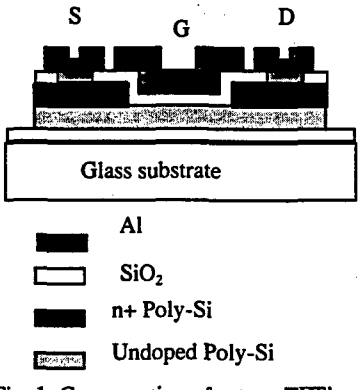



Fig. 1. Cross-section of n-type TFT's

The density of states (DOS) in the gap of the polysilicon is calculated from the tranconductance of the TFT's. Indeed, this transconductance is controlled by a band bending associated with a volume average DOS. By using an incremental method [3] to determine the relation between the gate bias and the band-bending at the oxide-channel interface, the DOS at an energy level is then calculated [4].

## 3. Results

Figure 2 shows the known improvement of the transfer characteristics after hydrogenation. The improvement is due, as known [5], to the passivation of the in-gap polysilicon defects by hydrogen. Hydrogen is known to passivate the silicon dangling bonds which are present in the grain boundaries of the polysilicon and also inside the grains of the solid phase crystallized polysilicon. The subsequent improvement of the electron transport properties of the silicon material induces a slight increase of the channel off-current at low reverse gate bias. The usual increase of the off-current at high reverse biases, due the defect induced field effect enhancement, is however reduced. In the on-regime, the reduced defect density induces a decrease of the subtreshold slope and a slight increase of the on-current. All these hydrogen passivation effects on the TFT's performances are known [2] and are not further considered. Particular attention is given in this paper to the degradation of the TFT's performances under gate bias stress and to the comparison between the reliability and then the stability of hydrogenated and unhydrogenated TFT's. Results of this stress study are then presented.

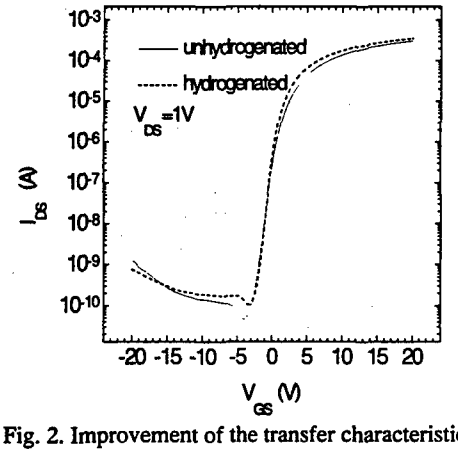



Fig. 2. Improvement of the transfer characteristics after hydrogenation

Figure 3 (respectively Figure 4) compares the behaviours of the threshold voltage shift $\Delta V_T$ (a), the subthreshold slope variation $\Delta S$ (b) of the two types of TFT's during negative (respectively positive) bias gate stress.

For negative stress the behaviours of the two types of TFT's are similar, the variations of the parameters are however weaker for hydrogenated TFT's. A positive shift of the threshold voltage is observed. The subtreshold slope increases and stabilizes. The stabilization occurs sooner in hydrogenated TFT's. This increase of S may be considred as an indication of a greater defect density.

For positive stress, the magnitudes of $\Delta V_T$ are equal for the two types of TFT's. The observed shift of the threshold voltage is negative. S increases and stabilizes for hydrogenated TFT's and is constant for unhydrogenated TFT's. Then, the defect density may

not depend on the positive stress in unhydrogenated TFT's.

## 4. Discussion

To discuss these results, it may be useful to give an overview of gate stress effects previously observed on hydrogenated amorphous and polycrystalline silicon TFT's. Mobile ions, such as $Na^+$ [6] or $OH^-$, $H^+$ water related ions [7], induce positive shift of the threshold voltage with negative stress and negative shift with positive stress, without any variation of the subthreshold slope or the transconductance. Trapping of carriers flowing at high gate bias into the oxide [8, 9] induces positive shift of the threshold voltage with positive stress and negative shift with negative stress, without any variation of the subthreshold slope. Carrier-induced states creation into the gap of the silicon was also involved to explain stress effects on hydrogenated amorphous silicon TFT's [10], which may be considered as normal, but also on hydrogenated polycrystalline silicon TFT's [11]. Creation of states is induced by positive and negative stress. It leads to positive (negative) shift of $V_T$ for n-type (p-type) TFT's.

States created in the upper (lower) part of the gap induce an increase of the n-type (p-type) subthreshold slope. The behavior of $\Delta V_T$ shows a power-law time dependence [10].

In n-channel c-Si MOSFET's, only the negative stress induces a shift of the threshlod voltage and an increase of S. This effect is mainly due to interface state creation and it is known at negative bias stress [12].

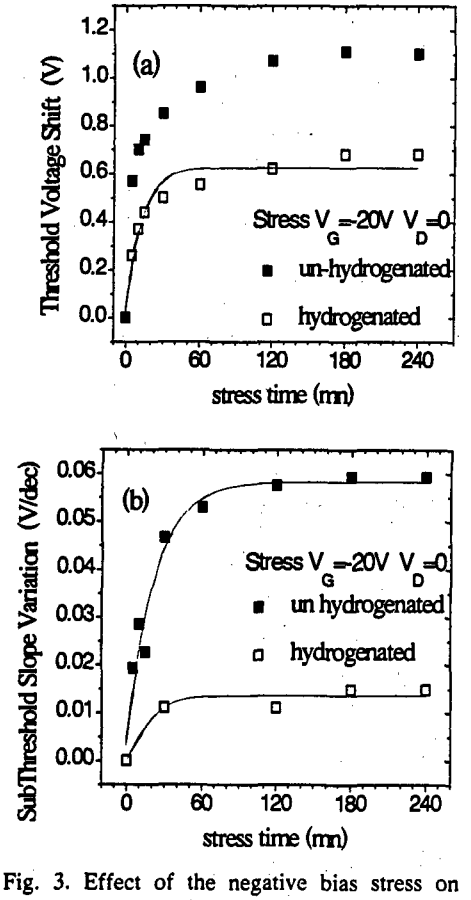



Fig. 3. Effect of the negative bias stress on the threshold voltage (a) and the subthreshold slope (b)

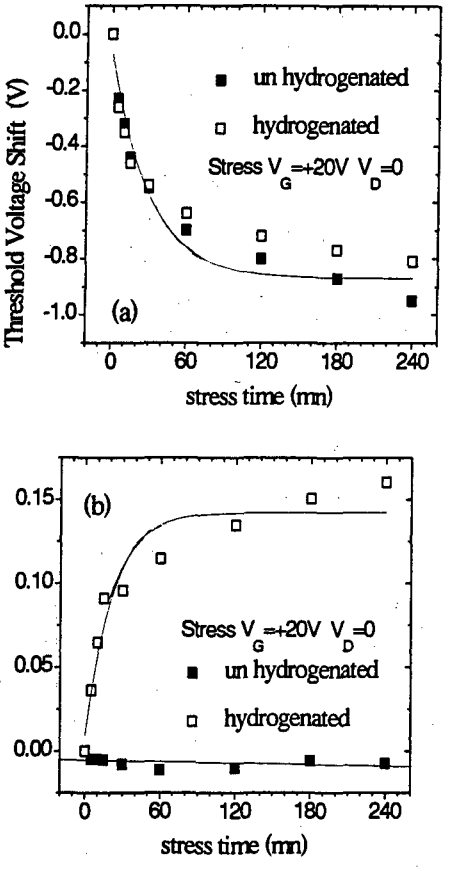



Fig. 4. Effect of the positive bias stress on the threshold voltage (a) and the subthreshold slope (b)

In the present study, the two types of TFT's show a positive $\Delta V_T$ for negative stress and negative one for positive stress. This sign of $\Delta V_T$ may be explained by the mobile ions effect if the subthreshold slope S does not vary. An increase of S is shown however in Figure 3b for the two types of TFT's and in Figure 4b for hydrogenated TFT's. This increase of S might involve, for example, a defect creation but $\Delta V_T$ is negative for positive stress.

Present results cannot be then explained by only one of the previous effects. More than one effect must be involved to explain the behaviours.

### 4.1. Hydrogenated TFT's

First, we can deal with hydrogenated TFT's. In this TFT's, the sign of $\Delta V_T$ is opposite to that of the stress. S increases for both stresses. The variation of $V_T$ is then due to a different effect from that inducing S changes. The sign of $\Delta V_T$ may be explained predominantly by the mobile ions effect. Water related species may not be involved here because of the high temperature used during the densification of the oxide. Sodium ions may be mobile however in present APCVD oxide particularly because it was not phosphorus doped. Phosphorus doping of oxide is known to reduce the mobility of sodium ions [6].

Degradation of the subthreshold slope S must then be explained from another effect. An increase of S may be explained by a state creation at the insulator-channel interface and/or in the channel material. Figure 5 shows the Density of States (DOS) profile in the upper part of the gap (above the Fermi level defined at the flat band equilibrium) before and after positive (Figure 5.a) or negative (Figure 5.b) stress. This DOS profile is determined from the transfer characteristics of hydrogenated TFT's. The behaviours show an increase of the density of deep states after the two types of stress.

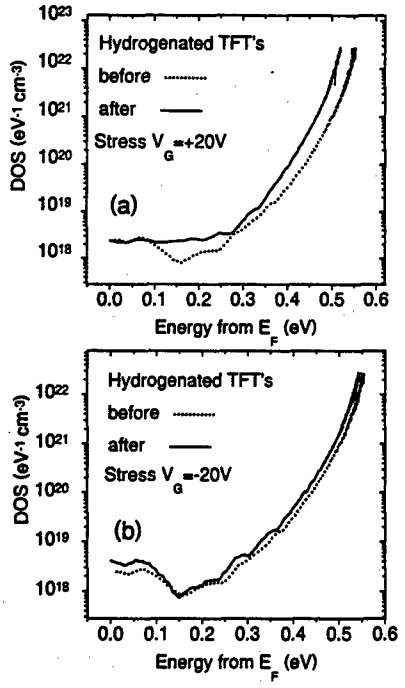



Fig. 5 : Effect of positive (a) and negative (b) stress for hydrogented TFT's

States creation is usually used to explain the degradation of hydrogenated amorphous and hydrogenated polycrystalline silicon TFT's. According to the model currently used [13], the weak Si-Si bonds, present in amorphous silicon or in strained and disordered regions of polycrystalline silicon, likely to break under electron or hole accumulation. This breaking induces pairs of dangling bonds defects. To prevent reforming of the weak bonds, the model needs to stabilize the dangling bonds by a diffusive motion of hydrogen.

This model may be involved to explain the increase of S in hydrogenated TFT's, more important when the stress bias is positive *i.e.* under electron accumulation. It is due to creation of states in the upper part of the band-gap in the case of n-type TFT's.

### 4.2. Un-Hydrogenated TFT's

In this TFT's, the sign of $\Delta V_T$ is also opposite to that of the stress. However S increases only in the negative stress case. No variation of S is observed with positive stress. In the last case, the origin of the stress effect is then predominantly due to the mobile ions. It seems that positive stress does not induce defect creation in unhydrogenated TFT's. Defect creation is induced in these TFT's only with negative stress when S increases.

Figure 6 shows the Density of States (DOS) profile in the upper part of the gap before and after positive (Figure 6.a) or negative (Figure 6.b) stress. This DOS profile is determined from the transfer characteristics of unhydrogenated TFT's. The behaviours show an increase of the DOS after the negative stress but the positive stress does not induce any DOS variation.

This behaviour of the unhydrogenated TFT's under positive or negative gate bias stress is typical of n-channel c-Si MOSFET's [12] where the state creation at the insulator-channel interface is involved with negative stress.

To understand the state creation in the present TFT's, weak bond breaking may be involved. The weak bonds are known to be present in disordered structure as grain boundaries, bulk of the grains in low temperature crystallized silicon, or $Si\text{-}SiO_2$ interface. Here stabilization of the new structure equilibrium by hydrogen motion cannot be involved however since TFT's are unhydrogenated. Possible explanation is that the bond breaking induces large scale structural changes which are favoured in regions of large disorder as unhydrogenated disordered regions in silicon. New structural equilibrium with new energy distribution of states set up. Such structural changes have been highlighted by Hata *et..al.*[14]. In the hydrogenated amorphous silicon, only bond breaking with creation of dangling bonds close to an atomic hydrogen and without, or with low, large scale structural changes, may occurs.

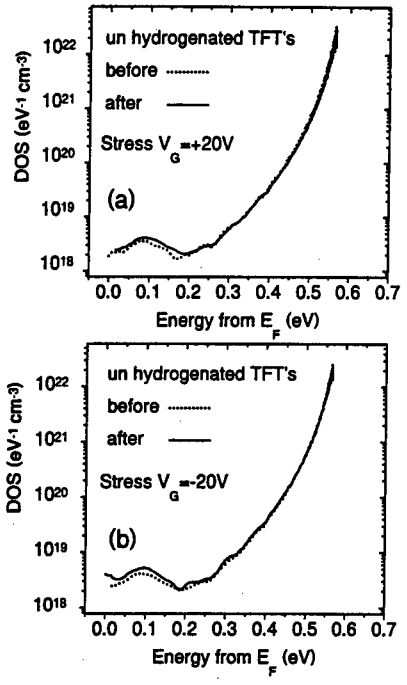



Fig 6: Effect of positive (a) and negative (b) stress for unhydrogenated TFT's

## 5. Conclusion

The lack of reliability of hydrogenated polysilicon TFT's is often related to the hydrogen behaviour in the material. In this way the stability of new high performance unhydrogenated polysilicon TFT's is questionnable.

In this work the reliability of two types of polysilicon TFT's originated from the same process was studied. One type of TFT's was unhydrogenated and the other was submitted to a Radio-Frequency hydrogen plasma. Then any difference in the reliability of these TFT's is only attributed to the hydrogenation step.

A reduced negative stress effect in hydrogenated TFT's is observed. A degradation occurs however with positive stress. This behaviour is typical of hydrogenated amorphous silicon or hydrogenated polysilicon TFT's.

More supprising is the behaviour of unhydrogenated TFT's characterized by an effect of negative stress and no effect of positive stress on S. Their behaviour is however similar to that of single-crystalline MOSFET's under negative gate bias stress. This lack of stability is explained from the structural changes will occur in disordered regions of silicon. The more important effect in polysilicon TFT's is only due to the importance of these disordered regions in the material. In polysilicon material, these regions are the grain boundaries, the bulk of the grains in low temperature crystallized silicon and the Si-$SiO_2$ interface. In single crystalline silicon, only the Si-$SiO_2$ interface may be considered as a disordered region.

## References

[1] Mourgues K., Raoult F., Pichon L., Mohammed-Brahim T., Briand D. and Bonnaud O. Materials Research Society Symp. Proc. 471 (1997) 155
[2] Petinot F. PhD Thesis, University of Paris VI, France, 1998
[3] Suzuki T., Osaka Y. and Hirose M. Jap. J. Appl. Phys. 21 (1982) L159
[4] Fortunato G. and Migliorato P. Appl. Phys. Lett. 49 (1986) 1025
[5] Johnson N., Biegelsen D., Meyer M. Appl. Phys. Lett. 40 (1982) 10
[6] Young N.D., Gill A. and Clarence I.R. J. Appl. Phys. 66 (1989) 187
[7] Young N.D. and Gill A. Semicond. Sci. Technol. 7 (1992) 1103
[8] Powell M.J., van Berkel C. and Hughes J.R. Appl. Phys. Lett. 54 (1989) 1323
[9] Young N.D. and Gill A. Microelectronic Eng. 19 (1992) 105
[10] Powell M.J., van Berkel C. and Franklin A.R. Phys. Rev. B45 (1992) 4160
[11] Young N.D. and Ayres J.R. IEEE Trans. Electron Devices ED-42 (1995) 1623
[12] Jeppson K.O. and Svensson C.M. J. Appl. Phys. 48 (1977) 2004
[13] Jackson W.B. Phys. Rev. B43 (1990) 1059
[14] Hata N., Kamei T., Okamoto H. and Matsuda A. Materials Research Society Symp. Proc. (1997)

PERGAMON

Microelectronics Reliability 38 (1998) 1155–1164

MICROELECTRONICS RELIABILITY

# Use of wafer maps in integrated circuit manufacturing

C.K. Hansen[a], P. Thyregod[b]

[a] *Department of Mathematics, Eastern Washington University, Cheney, WA 99004, USA*
[b] *Institute of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark*

**Abstract**

A wafer map identifies the locations of defective integrated circuits (chips) on a silicon wafer and provides important spatial information. The wafer yield is a useful measure of the process quality, but other features are necessary to account for. Careful statistical analysis addressing the spatial information in the wafer map is necessary in order to monitor the quality of the manufacturing process, and identify/eliminate fault sources with assignable causes. We discuss simple descriptive analyses of wafer map data, as well as formal statistical methods, based on three different models that account for different spatial patterns. In particular, the models support the observed phenomenon that the faults are distributed non-uniformly across the wafer, by allowing the fault probability to vary across the wafer, and allowing faults at adjacent locations to be statistically dependent.
© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Analysis of fault data for integrated circuits manufactured on silicon wafers have been a key focus in engineering and statistical literature during the past two decades or more. The simplest and most commonly used measure of process quality is the *wafer yield*, usually defined as the fraction of chips passing initial electrical and visual wafer level tests, and continuing on to packaging and subsequent final testing. The statistical properties of the yield statistic have been discussed extensively under various model assumptions, see e.g. [1,2,3]. Recent papers have introduced more comprehensive measures of wafer quality that address the spatial information contained in the location of the defective chips [4,5,6,7,8]. The use of co-fabricated test structures for cost-effective process control are discussed in [9,10,11].

A wafer map provides a simple graphical representation of the spatial locations of the defective chips on the wafer. Certain spatial patterns can be directly assigned to specific fault sources which, if properly identified, can be partially or completely eliminated. Photo-mask misalignment, over-etching, handling problems etc. usually result in spatial patterns that deviate from those observed for faults related to clean room conditions. Often conclusions are reached on the basis of informal "eye-ball" analysis rather than formal mathematical/statistical analysis. In this paper we provide an overview of informal descriptive analyses, as well as formal inferential statistical methods applied to wafer map data. The methods are easily applied by engineers concerned with integrated circuit fabrication process control and quality assurance, and are introduced at the user level with mathematical details kept to a minimum.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00127-9

### 1.1 Mathematical notation

| | |
|---|---|
| $(x, y)$ | Chip location. |
| $n$ | Number of chips on the wafer. |
| $N$ | Number of wafers in the lot. |
| $w$ | Wafer number within lot. |
| $X_w$ | Number of defective chips on the wafer $w$. |
| $Z_w$ | Number of working chips on the wafer $w$. |
| $X_{(x,y)}$ | Number of defective chips within lot at location $(x, y)$. |
| $Z_{(x,y)}$ | Number of working chips within lot at location $(x, y)$. |
| $p(x, y)$ | Fault probability for chips at location $(x, y)$. |
| $\widehat{p}(x, y)$ | Estimated fault probability. |
| $Y_w$ | Sample yield of wafer $w$. |
| $y$ | Expected wafer yield $y = E[Y]$. |
| $D_{(x,y)}$ | Defect density at location $(x, y)$. |
| $A$ | Critical area. |
| $X \sim B(n, p)$ | $X$ is a binomial random variable with number parameter $n$ and probability parameter $p$. |
| $X \sim N(\mu, \sigma)$ | $X$ is a normal random variable with mean $\mu$ and standard deviation $\sigma$. |
| HBP | Homogeneous Bernoulli process. |
| NHBP | Nonhomogeneous Bernoulli process. |

## 2. Mathematical modeling

In this study we consider binary fault data observed at the wafer level, often measured based on electrical wafer probe testing. Thus, each chip is classified as either defective (0) or working (1). A more detailed characterization is possible, e.g. by classifying faults according to the type of malfunction, or quantitatively, on the basis of some parameter required to remain within specified limits. Wafer maps based on such measures have been discussed in the literature [2,3], but will not be considered here.

We assume all chip faults to be generated as the result of some type of defect that can be described using a probability model. This does not impose any serious restriction on the type of patterns that can be observed, only that by assumption each chip is characterized by a fault probability, $0 \leq p \leq 1$,

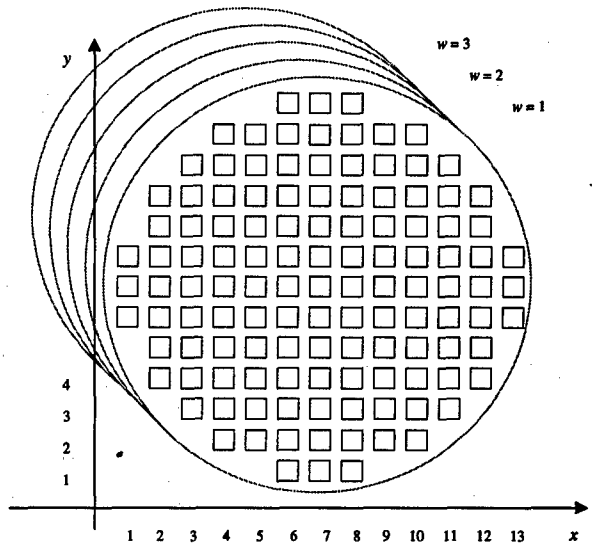



Figure 1: The $(x, y)$ coordinate system used to identify chip locations on a wafer map.

representing the probability that the chip will be defective when considered individually. By allowing fault probabilities to vary with location, and by allowing chip faults on the same wafer to be statistically dependent, we arrive at a fairly general model that can account for commonly observed patterns, such as a tendency that faults occur at a higher rate near the edge of the wafer compared to the inner region [9,11], and a tendency that chip faults occur in "clusters" [2-7].

We introduce three models that differ in the assumptions regarding fault probability uniformity (homogeneity), and dependence of faults at adjacent chip locations. The main focus will be on estimation of characteristic parameters for these models, and statistical tests that serve to distinguish the models. We also discuss a nonparametric approach that is not based on any specific model assumptions. Nonparametric analysis is useful as an initial guide in selecting an appropriate parametric model.

We consider a series of $N$ similar wafers usually representing wafers produced as part of the same lot. Each wafer consists of $n$ identical chips. The location of each chip is identified by its wafer number $w$ and its location within the wafer represented by its $(x, y)$ coordinates as illustrated in Fig. 1.

Each chip is characterized using a binary random

variable $C_w(x,y)$ as follows

$$C_w(x,y) = \begin{cases} 0 & \text{if chip is defective} \\ 1 & \text{if chip is working} \end{cases} \quad (1)$$

where the subscript $w$ identifies the wafer number. Correspondingly, we identify a working chip by an unfilled square in the wafer map, and a defective chip by a filled square. Unfortunately, there is not consensus about this notation in the literature. The total number of defective and working chips on wafer $w$ are denoted $X_w$ and $Z_w$ respectively,

$$X_w = n - \sum_{(x,y)} C_w(x,y) \quad (2)$$

$$Z_w = \sum_{(x,y)} C_w(x,y) \quad (3)$$

where the sum is extended over all existing locations $(x,y)$ on the wafer. In some cases we are interested in the total number of defective (and working) chips observed at a given wafer location $(x,y)$ which we denote $X_{(x,y)}$ and $Z_{(x,y)}$,

$$X_{(x,y)} = N - \sum_{w} C_w(x,y) \quad (4)$$

$$Z_{(x,y)} = \sum_{w} C_w(x,y) \quad (5)$$

where the sum is extended over all wafers.

### 2.1 The homogeneous Bernoulli process (HBP model)

The simplest possible model is obtained by assuming that all chips have the same fault probability,

$$p = P(C_w(x,y) = 0) \quad (6)$$

and that all chip faults are independent, in which case we arrive at a *homogeneous Bernoulli process* (HBP). The number of defective chips, as well as the number of working chips, are binomial random variables,

$$X_w \sim B(n,p) \quad (7)$$

$$Z_w \sim B(n,1-p) \quad (8)$$

$$X_{(x,y)} \sim B(N,p) \quad (9)$$

$$Z_{(x,y)} \sim B(N,1-p) \quad (10)$$

A simple diagnostic test is obtained by comparing the observed frequency distribution for either of the four statistics $X_w, Z_w, X_{(x,y)}, Z_{(x,y)}$ with its "theoretical" binomial distribution, as proposed by Kaempf [12]. The HBP model has a theoretical justification, when chip faults are produced as result of micro defects randomly distributed across the wafer. A Poisson process with constant intensity, referred to as the *defect density* $D$, is the appropriate mathematical model for the distribution of such defects. The theoretical relationship between defect density, fault probability and yield can be summarized as

$$p = 1 - y = 1 - e^{-DA} \quad (11)$$

where the parameter $A$ is referred to as the critical chip area. Defects of different size and type are combined using an average defect density and a corresponding combined critical area, see e.g. [10,11].

Control charts for monitoring the observed wafer yield based on the HBP model can be constructed by observing that the yield

$$Y_w = \frac{Z_w}{n} \quad (12)$$

is approximately normally distributed

$$Y_w \sim N(y,\sigma) \quad (13)$$

where $y$ is the target yield and

$$\sigma = \sqrt{\frac{y(1-y)}{n}} \quad (14)$$

Control charts based on more general model assumptions are discussed in [4,7].

### 2.2 The nonhomogeneous Bernoulli process (NHBP model)

The HBP model is not adequate in modeling situations where chip faults occur as result of defect mechanisms that systematically "favor" certain locations or regions of the wafer. A simple generalization of the model, referred to as the *nonhomogeneous Bernoulli process* (NHBP), is obtained by assuming that chip faults are independent, while allowing the fault probability to vary with location $(x,y)$. Studies based on actual data have suggested that the fault probability near the edge (outer region) of the wafer is higher than the defect density near the center (inner region), see [9,12]. A model that specifically takes this effect into account is the following

$$p(x,y) = p_0 + (p_1 - p_0)\frac{((x-x_0)^2 + (y-y_0)^2)}{R^2} \quad (15)$$

where $p_0$ is the fault probability near the center of the wafer, $p_1 > p_0$ the fault probability near the edge of the wafer, $(x_0, y_0)$ the center and $R$ the radius of the wafer. A more general model is

$$p(x,y) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 y + \beta_4 y^2 + \beta_5 xy \quad (16)$$

Clearly, the model given by Eq. 15 is a special case of Eq. 16. The first model has the advantage of having only two free parameters, both with simple interpretation. The parameters in Eq. 16 are estimated easily using the method of least squares as described later, but are more difficult to interpret. In general $p(x,y)$ can be any linear or nonlinear function of the location $(x,y)$.

### *2.3 Dependent chip faults (cluster model)*

A fault probability that varies from location to location can, to some extent, explain observed "clusters" of faults, i.e. groups of defective chips confined to the same region of the wafer. However, the NHBP model assumes a similar type of pattern for all wafers in the same lot. Often spatial clustering is observed on a given wafer in a pattern that is not reproduced on other wafers in the lot. This type of spatial dependence may be induced by "large" defects that affect several chips on the same wafer. There are several different ways to model chip to chip dependence mathematically. One possibility, discussed in [11], is to consider the defect density as a sum of two contributions,

$$D_{(x,y)} = D_0 + \xi_{(x,y)} \quad (17)$$

where $D_0$ is a constant (the deterministic contribution), and $\xi_{ij}$ a random variable (the random contribution). An "autoregressive" model for $\xi$ adapted from [13] is

$$\xi_{(x,y)} = \frac{a}{4} \cdot (\xi_{(x-1,y)} + \xi_{(x+1,y)} + \xi_{(x,y-1)} + \xi_{(x,y+1)}) + \epsilon_{(x,y)}$$
$$\epsilon_{(x,y)} \sim N(0,\sigma) \quad (18)$$

$(-1 < a < 1)$ which can be viewed as a generalization of the autoregressive time series defined on a discrete one-dimensional time scale $t$,

$$\xi_t = a\xi_{t-1} + \epsilon_t \quad (19)$$

The location-specific fault probability is then

$$p(x,y) = 1 - e^{-D_{(x,y)}A}$$

The NHBP model and the cluster model can be combined by replacing the constant $D_0$ with some deterministic function of the location $(x,y)$. Furthermore, the cluster model reduces to the HBP model when the parameter $\sigma$ is equal to zero.

A substantial complication when going from one- to two-dimensional autoregressive models is the lack of chronological order between the observations. In simulation studies, it introduces the complication that in order to generate the value of $\xi_{(x,y)}$, the value of (among others) $\xi_{(x-1,y)}$ must be known, but in order to generate the value of $\xi_{(x-1,y)}$, the value of $\xi_{(x,y)}$ must be known. Therefore the generation of $\xi_{(x,y)}$ values must be done using an iterative procedure.

## 3. Statistical analysis

For the remainder of this paper we shall focus our attention on statistical methods applicable to wafer map data. In order to verify the methods' effectiveness in identifying specific spatial patterns, we used simulation to generate three data sets corresponding to each of the models discussed in the previous section. In the examples we consider a lot consisting of 10 wafers each containing 121 chips. The simulated data are quite representative of spatial patterns that have been observed in real data. All data were analyzed without using any information about the model that generated the data. Calculations and graphics were produced using the symbolic software package Mathematica [14].

### *3.1 Nonparametric analysis*

A simple graphical representation of the entire lot of wafer maps can be constructed by plotting the estimated "local" fault probabilities (sample proportions)

$$\hat{p}(x,y) = \frac{X_{(x,y)}}{N} \quad (20)$$

as function of location $(x, y)$. This technique can be classified as *nonparametric*, since the fault probabilities are estimated with no assumptions made concerning the model generating the faults. When the number of wafers $w$ is small, however, the variance of these estimates is quite high, and it may be difficult to infer any spatial patterns from the plot. Compare the examples illustrated in Fig. 2-4. Therefore, some sort of "smoothing" of the raw plot is usually desirable. The most common nonparametric smoothing technique is based on averaging each local sample proportion with the sample proportions observed in the "neighborhood" of the location $(x, y)$,

$$\hat{p}(x,y)^* = \frac{1}{n^*} \sum_{(i,j)\in\mathcal{NB}} \hat{p}(i,j) \tag{21}$$

where the neighborhood $\mathcal{NB}$ is defined as

$$\mathcal{NB} = \{(i,j) \mid (i,j) \text{ is a valid location} \wedge x-1 \le j \le x+1 \wedge y-1 \le j \le y+1\} \tag{22}$$

and $n^*$ the number of locations included in $\mathcal{NB}$, typically 9, see also [6,7]. This type of plot may expose systematic variations in the fault probability, and may give some guideline as to which parametric model it may be appropriate to fit. Notice that the smoothed plot in Fig. 3 reveals a pattern not evident in the data shown in Fig. 2. This technique, however, is not useful for detecting spatial dependence, or clusters on individual wafers, since such effects tend to vanish when fault probabilities for several wafers are averaged. Thus, we can not clearly distinguish the smoothed probability plots shown in Fig. 2 and 4, even though the wafer map depicted in Fig. 4 suggests the presence of substantial clustering on individual wafer maps in this data set.

### *3.2 Fitting a nonhomogeneous Bernoulli process model using the least squares method*

A systematic spatial variation in the fault probability may be modeled using an NHBP model. As mentioned earlier, under this model faults are assumed to occur independently, whereas the fault probability varies with location according to some deterministic function $p(x, y)$. Suppose $p(x, y)$ contains a set of (unknown) parameters $\theta = (\theta_1, \theta_2 \ldots)$, then $\theta$ can be estimated using the ordinary least

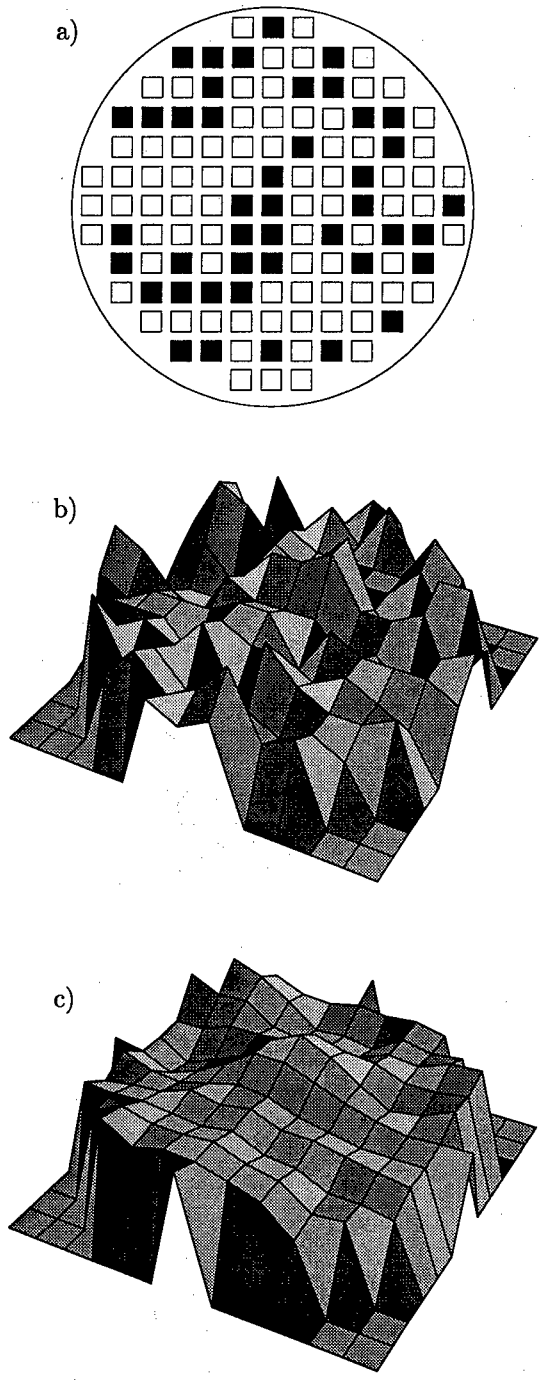
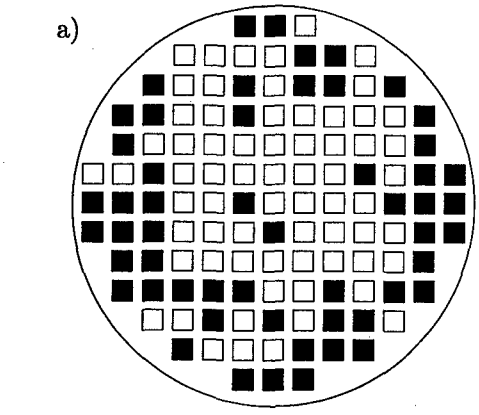
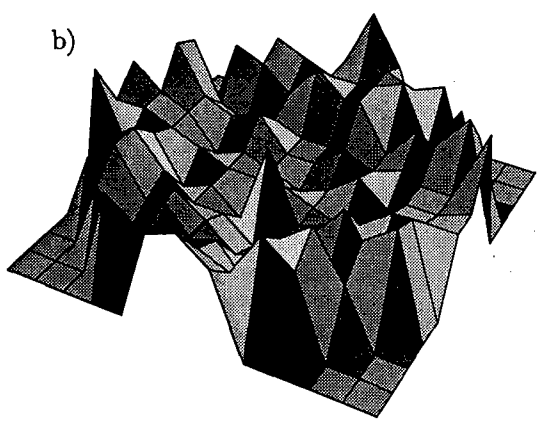



Figure 2: Simulated data from an HBP model. a) Sample wafer map (a filled square represents a defective chip). b) Raw fault probability plot based on 10 wafers. c) Smoothed fault probability plot.



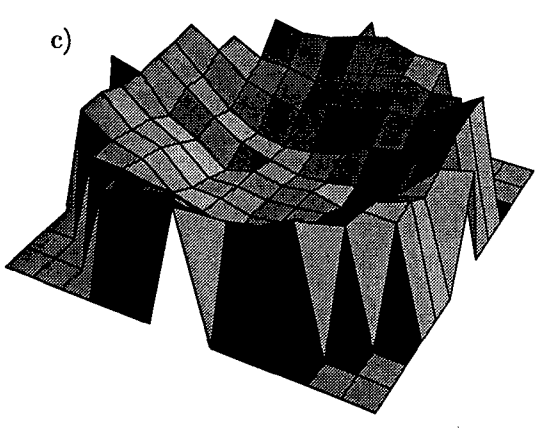
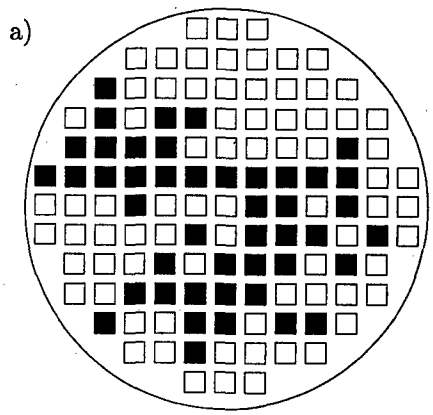
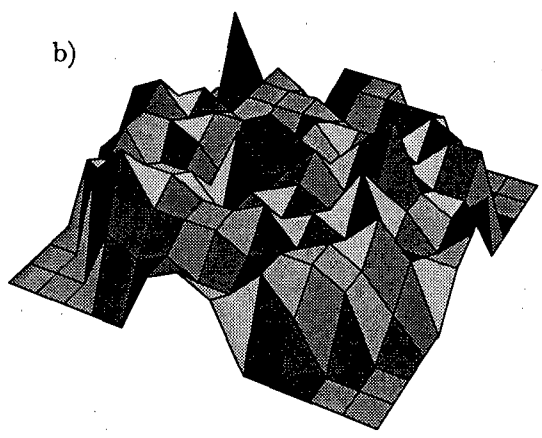




Figure 3: Simulated data from an NHBP model. a) Sample wafer map (a filled square represents a defective chip). b) Raw fault probability plot based on 10 wafers. c) Smoothed fault probability plot.



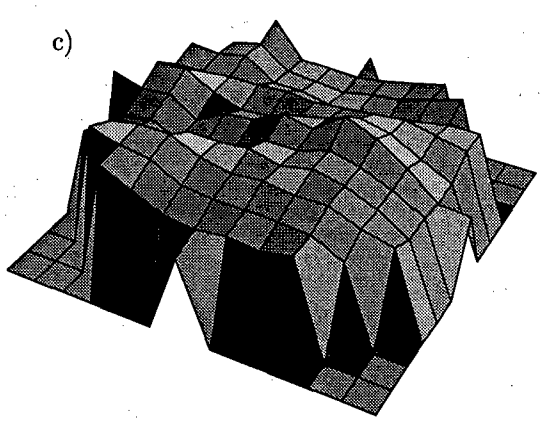




Figure 4: Simulated data from a cluster model. a) Sample wafer map (a filled square represents a defective chip). b) Raw fault probability plot based on 10 wafers. c) Smoothed fault probability plot.

Table 1: Least squares estimates of the parameters in the fault probability function $p(x,y) = p_0 + (p_1 - p_0)\left((x-x_0)^2 + (y-y_0)^2\right)/R^2$. Parameter estimates identified by an asterix (*) are significantly different from zero at a 5% significance level or lower.

| | HBP data | NHBP data | Cluster data |
|---|---|---|---|
| $\widehat{p_0}$ | 0.3862* | 0.1167* | 0.3496* |
| $\widehat{p_1} - \widehat{p_0}$ | 0.0027 | 0.5446* | -0.0571 |

Table 2: Least squares estimates of the parameters in the fault probability function $p(x,y) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 y + \beta_4 y^2 + \beta_5 xy$. Parameter estimates identified by an asterix (*) are significantly different from zero at a 5% significance level or lower.

| | HBP data | NHBP data | Cluster data |
|---|---|---|---|
| $\widehat{\beta_0}$ | 0.5083* | 1.6404* | 0.3469* |
| $\widehat{\beta_1}$ | 0.0014 | -0.2190* | 0.0084 |
| $\widehat{\beta_2}$ | -0.0011 | 0.0155* | -0.0016 |
| $\widehat{\beta_3}$ | -0.0286 | -0.2159* | 0.0026 |
| $\widehat{\beta_4}$ | 0.0012 | 0.0148* | -0.0016 |
| $\widehat{\beta_5}$ | 0.0012 | 0.0008 | 0.0017 |

squares method as the value of $\theta$ that minimizes the sum of squared errors

$$SSE = \sum_{(x,y)} (\widehat{p}(x,y) - p(x,y))^2 \tag{23}$$

Models that are linear in their parameters, such as Eq. 15 and 16 are particularly attractive, because a closed form solution can easily be obtained, see Tables 1 and 2. Furthermore, for moderate sample sizes, the estimator is approximately normal, which allows construction of simple $z$-tests and/or confidence intervals to determine if individual parameters differ significantly from zero. Once a model has been fitted, it may be appropriate to assess the goodness of fit, e.g. by comparing the fitted fault probability plot, see Fig. 5 to the nonparametric plot of estimated fault probabilities, or by subjecting the residuals

$$\epsilon(x,y) = \widehat{p}(x,y) - p(x,y) \tag{24}$$

to a test for spatial dependence as discussed below.

In Eq. 23 all squared deviations are given equal weight, however, the uncertainty of $\widehat{p}$ assumes much higher values when the actual $p$ is close to 0.5 compared to when it is close to 0 or 1. A procedure that takes this into account is the maximum likelihood method, which involves finding the value of $\theta$ that minimizes the *residual deviance*

$$D(\widehat{p},p) = \sum N d(\widehat{p}(x,y), p(x,y)) \tag{25}$$

where

$$\begin{aligned} d(\widehat{p},p) &= 2\{\widehat{p} \times \ln(\widehat{p}/p) \\ &+ (1-\widehat{p}) \ln\left((1-\widehat{p})/(1-p)\right)\} \end{aligned} \tag{26}$$

is the unit deviance for the binomial distribution, see e.g. [15,16]. Also here the parameter estimates are approximately normal. The calculations must be done using an iterative procedure, however, some statistical packages including *S-plus* and *SAS* have procedures that can easily handle this task. These procedures also produce a statistic for testing the goodness of fit of the model.

### *3.3. Testing for spatial dependence using the log-odds ratio test*

The *log-odds ratio test* has been discussed extensively e.g. in [17] in the context of analysis of two-way contingency tables, i.e. in testing dependence between two binary variables. [5] demonstrates how this test can be used as a test for spatial dependence. Here the null-hypothesis is based on the HBP model, i.e. the distribution of the test statistic is derived assuming that faults occur independently with a constant probability. Under this scenario, consider two locations $(x,y)$ and $(x',y')$ that are adjacent to one another, i.e., belong to the same neighborhood. For the log-odds ratio we have

$$\begin{aligned} \ell = \log\Bigg(&\frac{P(C(x,y)=1, C(x',y')=1)}{P(C(x,y)=1, C(x',y')=0)} \times \\ &\frac{P(C(x,y)=0, C(x',y')=0)}{P(C(x,y)=0, C(x',y')=1)}\Bigg) = \\ \log\Bigg(&\frac{(1-p)\cdot(1-p)}{(1-p)\cdot p} \times \frac{p\cdot p}{p\cdot(1-p)}\Bigg) = 0 \end{aligned}$$

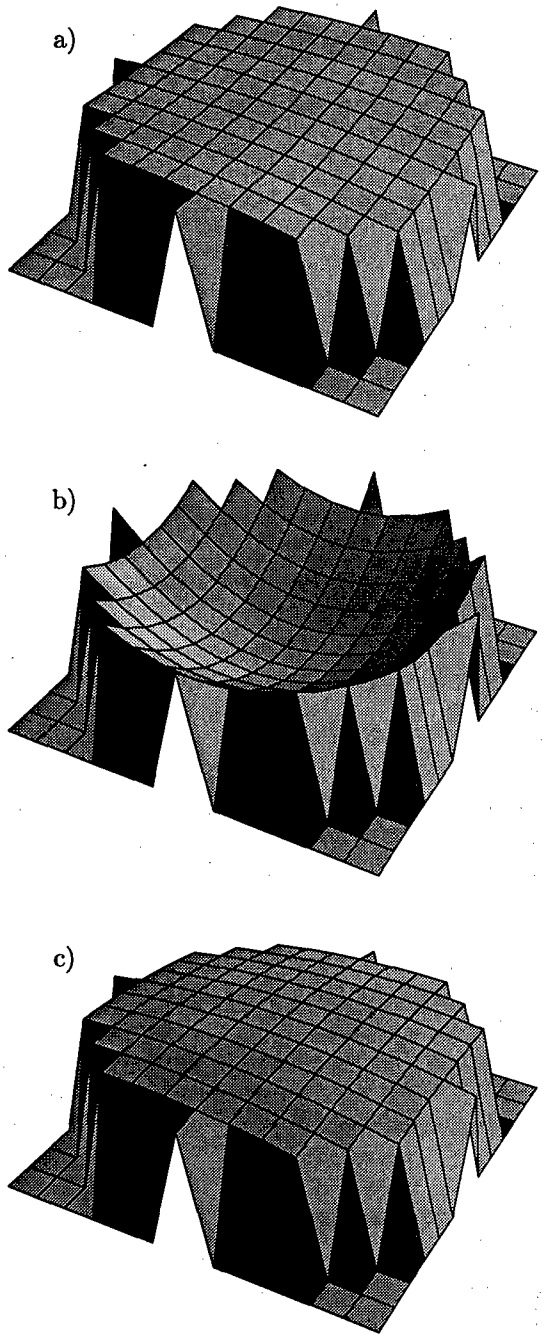



Figure 5: Fault probabilities fitted to a parametric model (Eq. 16) using the least squares method for data sets simulated under three different models a) HBP model b) NHBP model c) Cluster model. In b) the fault probability increases with radial distance from the center of the wafer, in a) and c) the fault probability is approximately constant.

Here the events in the numerators can be regarded as *attractions* between like outcomes represented by either two defective or two working chips, and the denominators as *repulsions* represented by one defective and one working chip adjacent to one another. An estimate of $\ell$ can be constructed using *join-count* statistics

$$\hat{\ell} = \log\left(\frac{JC_{ff} \cdot JC_{pp}}{(JC_{pf}/2)^2}\right) \tag{27}$$

The "fault-fault" join-count $JC_{ff}$ is found by counting the number of adjacent location pairs $(x, y)$ and $(x', y')$ for which both chips are defective, likewise, the "pass-pass" join-count $JC_{pp}$ is the number of adjacent location pairs $(x, y)$ and $(x', y')$ for which both chips are working. Since the ordering of $(x, y)$ and $(x', y')$ is arbitrary, we do not distinguish a "fault-pass" join from a "pass-fault" join, so we define $JC_{pf}$ as the number of adjacent location pairs $(x, y)$ and $(x', y')$ for which one chip is defective and one is working. This explains why $JC_{pf}$ is divided by 2 and then squared in Eq. 27. For small sample sizes [5] suggests adding a 0.5 correction to all join-count statistics to avoid any of these being zero.

For large sample sizes the estimator $\hat{\ell}$ is approximately normal with standard deviation

$$\sigma_{\hat{\ell}} \approx \sqrt{\frac{1}{JC_{ff}} + \frac{1}{JC_{pp}} + \frac{4}{JC_{pf}}} \tag{28}$$

see [17]. Therefore, a test statistic for testing the null hypothesis of an HBP model against a model with spatial dependence is

$$Z = \frac{\hat{\ell}}{\sigma_{\hat{\ell}}} \tag{29}$$

for which $p$-values can be found using a standard normal table. The results of log-odds ratio tests performed on the data sets generated using the HBP, NHBP and cluster models are summarized in Table 3. As expected, the test gives strong evidence of spatial dependence in the data generated from the cluster model, and no evidence of spatial dependence in the HBP data. The log-odds ratio test, however, is not efficient in distinguishing wafer maps observed under the NHBP and cluster models, and we observe that the NHBP data is misclassified as having spatial dependence. Therefore, it is recommended that the test be applied

Table 3: Results of Log-odds ratio test for spatial dependence. (*) based on a sample of 10 wafers. (**) based on only the individual wafers shown in Fig. 2-4.

| | HBP data | NHBP data | Cluster data |
|---|---|---|---|
| $J_{ff}$ (*) | 639 | 683 | 642 |
| (**) | 54 | 86 | 93 |
| $J_{fp}$ (*) | 1996 | 1848 | 1443 |
| (**) | 197 | 169 | 157 |
| $J_{pp}$ (*) | 1605 | 1709 | 2155 |
| (**) | 173 | 169 | 174 |
| Test statistic, $z$ (*) | 0.45 | 4.82 | 14.12 |
| (**) | -0.18 | 3.50 | 4.71 |
| $p$-value (*) | 0.65 | 1.45E-6 | $\approx 0$ |
| (**) | 0.86 | 4.63E-4 | 2.45E-6 |

only if there is no evidence of a systematic variation in the fault probability as determined by the methods discussed in the previous subsection. It is possible to have wafer map data that exhibit a combination of the characteristics of both the NHBP and cluster models, such that a systematic variation of the fault probability is observed commonly on all wafers, while there exist substantial clusters of faults on individual wafers that can not be explained by the NHBP model alone. While, as mentioned earlier, the NHBP and cluster models are easily combined mathematically, the statistical inference procedures applicable to such models are quite complicated, and will not be dealt with in this paper.

The log-odds ratio test can also be used in assessing the goodness of fit of a fitted NHBP model discussed in the previous subsection. When the model fit is good we should observe residuals (Eq. 24) that are randomly scattered around the fitted fault probability surface. A simple way to test the randomness of the residuals is by identifying residuals that are (at or) above the surface with a 1 and those below with a 0 in which case we obtain a set of binary spatial data, that can be subjected to the same analysis as regular wafer map data. Note that a log-odds ratio test performed on the residuals does not account for possible dependence of observed faults on the individual wafer maps. Table 4 shows the results of the log-odds ratio test performed using the residuals observed when fitting an NHBP model to each of the three data sets. All tests reveal no evidence of spatial dependence in the residuals.

Table 4: Results of Log-odds ratio test performed on residual maps obtained from fitting a parametric model (Eq. 15) to the estimated fault probabilities.

| | HBP data | NHBP data | Cluster data |
|---|---|---|---|
| $J_{ff}$ | 68 | 102 | 89 |
| $J_{fp}$ | 191 | 228 | 197 |
| $J_{pp}$ | 165 | 94 | 138 |
| Test statistic, $z$ | 1.01 | -1.56 | 1.20 |
| $p$-value | 0.31 | 0.12 | 0.23 |

## Conclusions

Careful statistical analysis of integrated circuit manufacturing data is essential when the objective is to identify assignable cause fault mechanisms that can be partially or completely eliminated with an improved product quality as a result. We have introduced three models that describe different spatial patterns commonly observed in a manufacturing process. The HBP model is an appropriate model in the case where faults tend to occur in a uniform and random fashion, with no particular spatial pattern in the observed faults. The NHBP model is advocated if a similar systematic variation in the fault probability is observed for all wafers in the lot. Finally, the cluster model is proposed as an alternative appropriate in the case where faults at adjacent chip locations appear to be statistically dependent, but with no evidence of a systematic pattern that can be traced from wafer to wafer. The methodology discussed in the paper is concerned with identifying spatial characteristics of manufacturing data by identifying the data with one of these three models.

The models introduced here serve only as first

order approximations to the many different patterns that may be observed in real world data, however, they often give a sufficient description of the most fundamental properties of interest for the purpose of identifying the mechanisms generating chip faults.

## Acknowledgement

This research was funded in part by a grant from the Northwest Institute of Advanced Study.

## References

1. Stapper C.H., Armstrong F.M. and Saji K. Integrated circuits yield statistics. Proceedings of the IEEE. Vol. 71 (1983) 453-470.
2. Tyagi A. and Bayoumi M.A. Defect clustering viewed through generalized Poisson distribution. IEEE Trans. Semiconductor Manufact. Vol. 5 (1992) 196-206.
3. Collica R. The effect of the number of defect mechanisms on fault clustering and its detection using yield model parameters. IEEE Trans. Semiconductor Manufact. Vol. 5 (1992) 189-195.
4. Friedman D.J. Some considerations in the use of quality control techniques in integrated circuit fabrication. International Statistical Review. Vol. 61 (1993) 97-107.
5. Taam W. and Hamada M. Detecting spatial effects from factorial experiments: An application from integrated circuit manufacturing. Technometrics. Vol. 35 (1993) 149-160.
6. Hansen M.H, Nair V. and Friedman D.J. Model-free estimation of defect clustering in integrated circuit fabrication. IEEE Trans. Semiconductor Manufact. Vol. 10 (1997), 344-359.
7. Hansen M.H, Nair V. and Friedman D.J. Monitoring wafer map data from integrated circuit fabrication processes for spatially clustered defects. Technometrics, Vol. 39 (1997) 241-253.
8. Stine B.E., Boning D.S. and Chung J.E. Analysis and decomposition of spatial variation in integrated circuit processes and devices. IEEE Trans. Semiconductor Manufact. Vol. 10 (1997) 24-41.
9. Amor S.B., Gregoris G., Riviere V., Touboul A., Stevenson J.L., Yeung P.S. and Hansen C.K. Test structures for reliability assurance in spacecraft microelectronics. In: Proceedings of the ESREF 93 Conference, Bordeaux, France, 1993, pp. 71-76.
10. Hansen C.K. and Thyregod P. Modelling and estimation of wafer yields and defect densities from microelectronics test structure data. Quality and Reliability Engineering International. Vol. 12 (1996) 9-17.
11. Hansen C.K. Effectiveness of yield-estimation and reliability-prediction based on wafer test-chip measurements. In: Proceedings of the Annual Reliability and Maintainability Symposium. Philadelphia, PA, 1997, pp 142-148.
12. Kaempf U. The binomial test: A simple tool to identify process problems. IEEE Trans. Semiconductor Manufact. Vol. 8 (1995) 160-166.
13. Ripley B.D. Spatial statistics. John Wiley, New York, 1981.
14. Wolfram S. The Mathematica book (3rd edn). Cambridge University Press, Cambridge, UK, 1996.
15. McCullagh P. and Nelder J.A. Generalized linear models (2nd edn). Chapman & Hall, London, 1989.
16. Jørgensen B. The theory of dispersion models. Chapman & Hall, New York, 1997.
17. Agresti A. Categorical data analysis, John Wiley, New York, 1990.

PERGAMON

Microelectronics Reliability 38 (1998) 1165–1169

MICROELECTRONICS RELIABILITY

# Field failure analysis on transmission data equipment due to lightning discharges

E. Mino Diaz[a], J. E. Vila Aresté[a]

[a]Telefónica Investigación y Desarrollo, Emilio Vargas 6, Madrid 28043, Spain

## Abstract

The telecommunication equipments suffer more damages, dues to lightning and power lines induction, than other type of appliances. The cause is that they have direct connections with wires or cables that are exposed to transients, and usually the affected components have direct connections with external lines, but sometimes the mechanism is more subtle and can affect circuits without contact with exposed connections.

This paper treats about failures on data interfaces circuits induced by voltage transients on the earth protection wire. It is based on a study of modems failures connected to remotes data terminals through RS-232 interfaces. Other data interface could also be damaged by this mechanism of failure (RS-422, Ethernet, Token Ring, etc.).

On this type of failure the failed circuits (drivers and receivers RS-232) are not located nearly the transient entry gate (telephone line and VAC inputs). They are in an interface between two equipments, and usually the only failed components are the drivers and receivers. © 1998 Elsevier Science Ltd. All rights reserved.

## 1.Field failure analysis

Several field failed equipments were collected and analysed they present two kinds of mechanisms and failure mode:

- About 30% of the failed equipments showed fractures on the driver/receiver epoxy body, internal severe damage and there were another failed integrated circuits.
  The power supply displays arcing among the earth protection, RS-232 common and 220 V traces. In some cases, due to the high current that ran through the RS-232 common trace, it was deformed.
- About 70% showed the drivers/receivers as the only failed components. It was not possible to observe an external damage over integrated circuit body, but some of them showed an internal damage that can be observed using a SEM (see Fig.1).
  The power supply didn't display arcing.

The cause of the first mechanism is a raising of the earth potential that cause arcing towards near traces acting upon all kinds of components.

Fig.2 shows a board with arcing amongst earth and GND and other traces. In this board the current through RS-232 common trace was so high that the trace was deformed as show Fig.3, and provokes the fracture in the driver (see Fig.4). Electrically the drivers and receivers presented severe damages.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00130-9

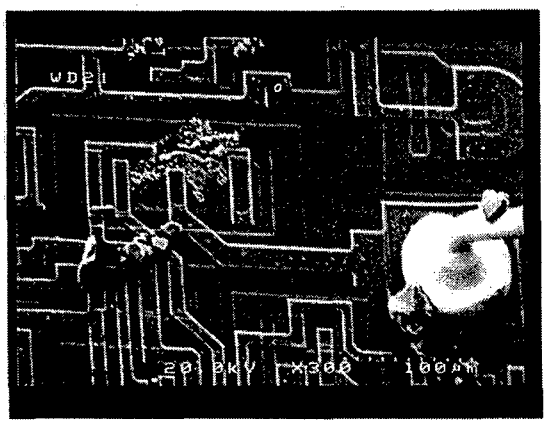


Fig.1. SEM view of a melted output MOS transistor in an RS-232 driver.

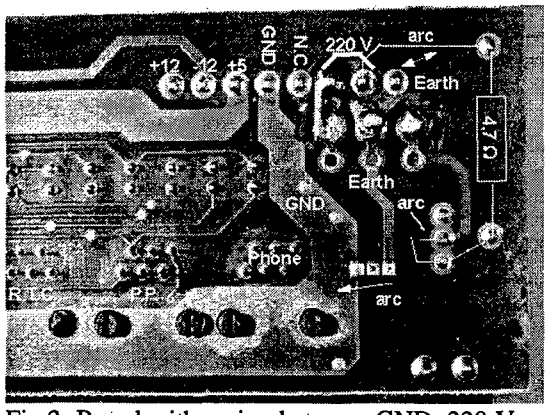



Fig.2. Board with arcing between GND, 220 V and earth traces

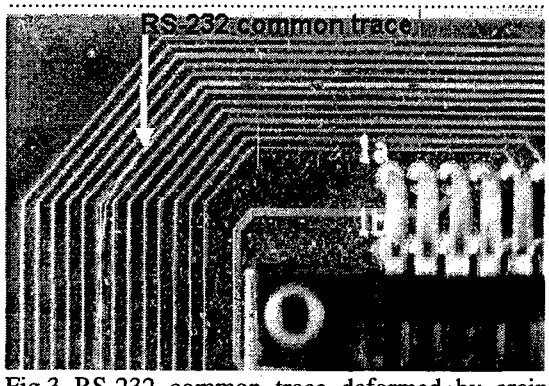



Fig.3 RS-232 common trace deformed by arcing current from earth protection connection.

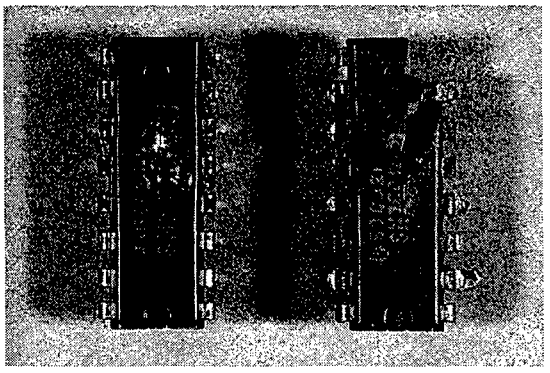


Fig.4. Fractures on the driver/receiver epoxy body

The drivers/receivers that failed with the second mechanism presented an anomalous electrical behaviour. Drivers had output voltages some differents of the typical values, + 12V or – 12V. Receivers had input resistances some differents of the typical 5 kΩ or 6 kΩ, depending of the component manufacturer.

*1.1 Drivers and receivers electrical characteristics*

The receivers can support some degree of overvoltage. Fig.5 shows the voltage-current curves of a receiver, the 75C1406, when a voltage sweep between -100 V and 100 V was applied, this sweep didn't provoke damage in the circuit. The bottom line is the input voltage-current characteristic (input resistance 5 kΩ). The top line is the output voltage (the output voltage change from 5 V to 0 V when the input voltage is 3 V).

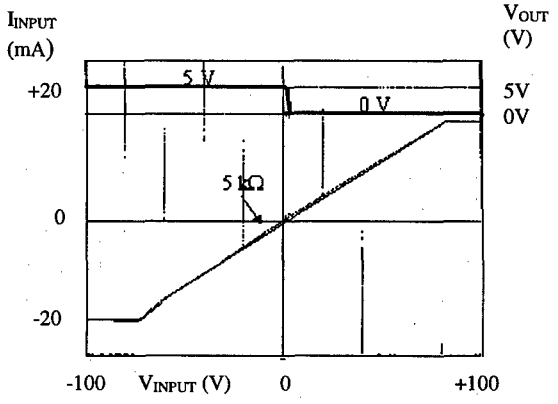



Fig.5 Receiver voltage-current curves

The driver also present some grade of protection, in fact the RS-232 standard state that all output or input terminals can be short-circuited. Fig.6

presents the output voltage-current curve of a driver, with an output voltage of 15V, when a voltage sweep between -25V and 25V is applied. It can be seen as between –12V and 17V the component limits the current at –12mA, and for voltages besides 17V the output resistance is 300Ω. But at –22V the component fall in breakdown.

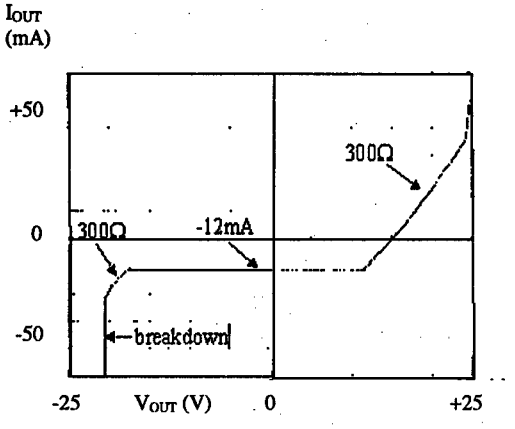



Fig.6 Driver current –voltage curve

The next paragraphs explain the coupling mechanism that provokes the second mode of failure (coupling of a transient on the earth protection wire).

## 2.Discharge coupling

Fig.7 shows the connections between a modem and a remote data terminal in the subscriber premises and their connection to the power supply line. It is important to observe that it includes an earth line connected to both equipments, moreover the RS-232 connection. The distance between both appliances could be as long as 20 meter.

Fig.8 shows the equivalent circuit with the most important components and connections. It includes one data line, the common wire of the RS-232 interface that connects the two ground wires of the DC powers supplies (COM1 and COM2) and the earth line.

The earth line has a direct connection between the COM 1 and the earth line in the PC power supply and a decoupling capacitor (C4) between earth wire (T2) and COM2. The equivalent circuit also shows the inductance of the lines and the receiver input impedance.

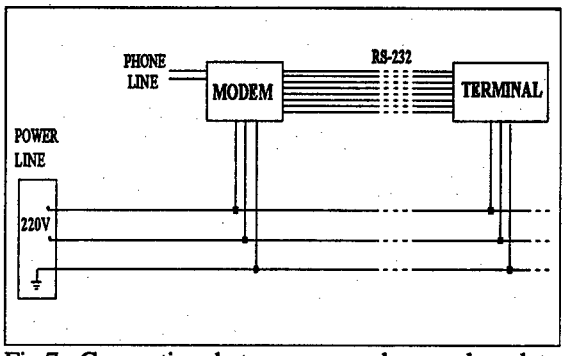



Fig.7. Connection between a modem and a data terminal

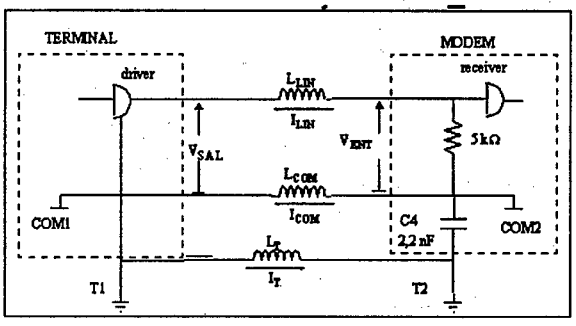



Fig.8. Equivalent circuit.

When a fast transient go through T1 and T2, some voltage is induced on $L_{COM}$ through C4. The voltage across the receiver of the modem is:

$$V_{ENT} = V_{SAL} + V_{COM} - V_{LIN}$$

As $L_{LIN}$ has a series resistance of 5-6 kΩ (the typical RS-232 receiver input resistance) the current trough $L_{LIN}$ is lower than the current trough $L_{COM}$ and a despicable voltage drop on $L_{LIN}$. Then the voltage induced on $L_{COM}$ is divided between the driver and receiver impedances:

$$V_{COM} = V_{ENT} - V_{SAL}$$

Then the interference transient voltage appears at the output of the driver and at the input of the receiver and they could be damaged.

## 3.Test and measurements

A transient voltage with a waveform 1.2/50 μs was applied to real equipments between the earth connections at the terminal and modem (T1 and T2) in order to try to reproduce the field failures in the laboratory.

The voltage obtained at the driver output was recorded. The result when a voltage of 1000 V was applied is shown in Fig.9. The waveform has a peak value of 70 V and the frequency is 333 kHz.

Several tests were performed at different transient voltages, obtaining similar waveforms. At a level of 1500 V the driver and/or the receiver were damaged in a similar way that the second type field failed components. Also it is important to note that they were the only one failed.

Voltages above 2500 V produced arcing, drivers and receivers were fractured and other components were affected as the first type of field failures. At voltages higher than 3500 V it was observed common RS-232 deformation, like the field failures (see Fig.3).

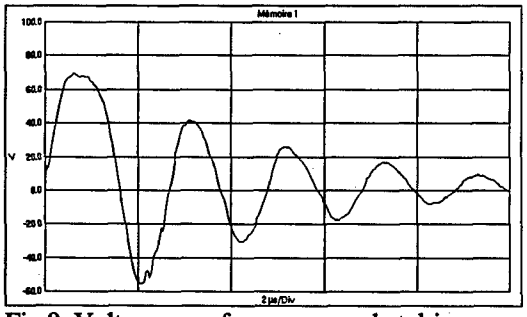



Fig.9. Voltage waveform measured at driver

## 4.Circuit simulation

It has been simulated the behaviour of the laboratory test circuit using SPICE. The circuit is shown in Fig.10. The circuit parameters are the following:

- 6 kΩ resistances replaced the driver and receiver.
- It is supposed a data line with 5m of cable, with an inductance of 1.6 μH/m (8 μH) and an ohmic value of 0.2 Ω/m (1 Ω).
- The transient generator had a waveform 1.2/50 μs and a series resistance of 2 Ω.
- The coils L4 and L5 correspond to the 220 V connection cables.
- C1 is the decoupling capacitor at the modem.

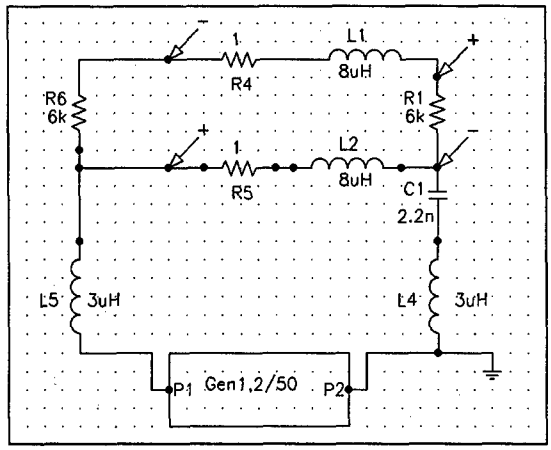



Fig.10. Simulated circuit

Fig. 11 shows the voltage waveform on each 6 kΩ resistance, when applied a discharge of 1000 V, with a waveform of 1.2/50 μs. The discharge induced a damped sinewave with a peak value of 104 V and a frequency of 835 kHz.

A resume of the most important peak voltages and currents values is the following:

- V (R1) = 104 V
- V (R6) = 104 V
- V (C1) = 1387 V
- V (L1) = 1.7 V
- V (L2) = 206 V
- I (L2) = 6.6 A
- I (L1) = 20 mA

It is important to note that the peak voltage value obtained in L1 is much lower than the peak voltage value in L2. That confirms the previous assumptions in the sense that the transient voltage coupled in the RS-232 common wire is divided between the driver and receiver.

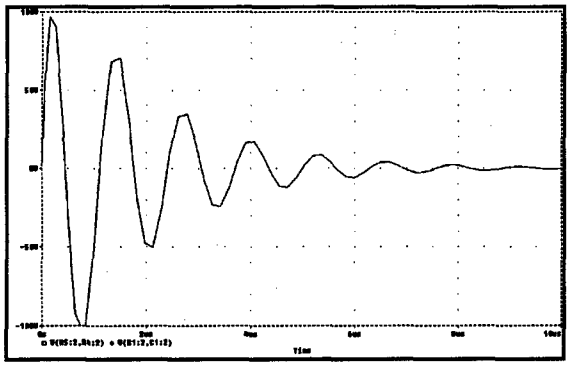


Fig.11. Voltage waveform at the simulated circuit

The waveform is very similar to the laboratory measured, although in the real measurements the peak value was 70 V and the frequency was 333 kHz. The differences among real and simulated waveforms are due to: the driver protections against electrostatic discharges (that limit the overvoltages), also in the real case the discharge affected to 22 data lines and in the SPICE model 1 line. Another factors are the mutual inductance between data lines and common wire, driver and receiver capacitance, etc.

## 5.Mathematical approach

It was developed the differential equation of the series RLC circuit, in order to evaluate the current and the voltage induced in the common wire. Supposing a step of E volts is applied and considering the real components values range; R (Ω), L (μH) and C (nF), then:

$$R^2 << 4.L/C$$

With this premise the current equation is very self-explanatory and simple:

It agrees with the simulated waveform and clearly it is possible to observe the influence of each part of the equation:

$$i = E \cdot \sqrt{\frac{C}{L}} \cdot e^{-\frac{R}{L} \cdot t} \cdot \sin \frac{1}{\sqrt{L \cdot C}} \cdot t$$

Amplitude | Damping | sinewave

- The sinewave is originated by the sin term and their frequency is the resonant frequency of the LC circuit.
- The signal envelope is originated by the exponential term. The inductance is charged by the first peak current and immediately follows their discharge, which has a time constant of L/R.
- The amplitude term that depends directly on the voltage step and the capacitance, and inversely with the inductance.

## 6.Conclusions

- The drivers and receivers are the usual failed components in the data transmission interfaces analysed, due to lightning discharges, and usually the only one failed.
- The failed components are not located nearly the transient entry gate (telephone line, VAC input), they are located in the interface between equipments.
- It is possible to reproduce the field failures applying discharges between both earth protection connection points. Depending on the voltage level there are two failure mechanisms
- At a level of 1500 V the driver and/or the receiver were the only one failed. The mechanism of failure was coupling of the transient on earth protection wire
- Voltages above 2500 V produced arcing, drivers and receivers were fractured and other components were affected. In this case the failure mechanism is arcing form the earth connections.
- It has been simulated the behaviour of the real circuit with discharges, obtaining a similar waveform to the measured at the drivers in the laboratory test, a damped sinewave. It allows to understand the coupling of the transient and to evaluate the applicable protection measures.
- A mathematical equation has been developed, that agrees with the real situation, and allows identify the different terms that generate the waveform and evaluate the influence of physical parameters.

PERGAMON

Microelectronics Reliability 38 (1998) 1171–1175

MICROELECTRONICS RELIABILITY

# Comparison between field reliability and new prediction methodology on avionics embedded electronics

P. Charpenel [a], P. Cavernes [b], V. Casanovas [a], J. Borowski [b], JM. Chopin [a]

[a] *AEROSPATIALE Aéronautique, 316 Route de Bayonne, M8621, 31060 Toulouse Cedex, France*
[b] *GIAT Industries, 155 Av. de Grande-Bretagne, 31052 Toulouse Cedex, France*

**Abstract :**

This paper presents a critical analysis of a deterministic reliability prediction approach previously described. Moreover, the proposed methodology is applied to an avionics embedded electronics equipment and the results are compared to field return data. A very good accordance between these predictive reliability data and field data is shown. The limitations of this methodology are analyzed. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

A previous paper [1] showed an other way to assess electronics reliability based on reliability tests performed by the parts manufacturers. Some critical points have been pointed out and have to be treated. This paper presents a critical analysis of this methodology and its application on an avionics embedded electronics equipment. Then, the results are compared to field return data and classical used methods.

## 2. Context

The identification and follow-up of our avionics is a prime necessity for our operational dependability policy. A systematic follow-up has been put in place, based on the research of the anomaly root-cause accounting for each equipment removal. Then, a set of indicators is created in order to control :

- reliability growth (according to DUANE model),
- relative repartition of failures observed during aircraft integration and during normal utilization by the airlines so that early failures are discriminated from wear-out one's,
- comparison for each component family between the operational and estimated failure rates.

Various prediction methods exist (MIL-HDBK-217 [2], CNET RDF [3], ...) but they are based on empirical failure rate models, developped from curve fits of field failure data. These data are limited in terms of number of failures in a given environment and determination of the actual cause of failure [4]. Moreover, future developments of these methods are doubtful and involve the development of new methodologies. Figure 1 shows field return results on an avionics computer, compared to the predictive failure rate.

## 3. Methodology description

Our electronics component policy is based on the use of components off the shelf (COTS) with certain levels of quality and reliability. We request all the necessary information from the manufacturer to assess his component technology, given the fact that he is alone to know his products, and then its

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00077-8

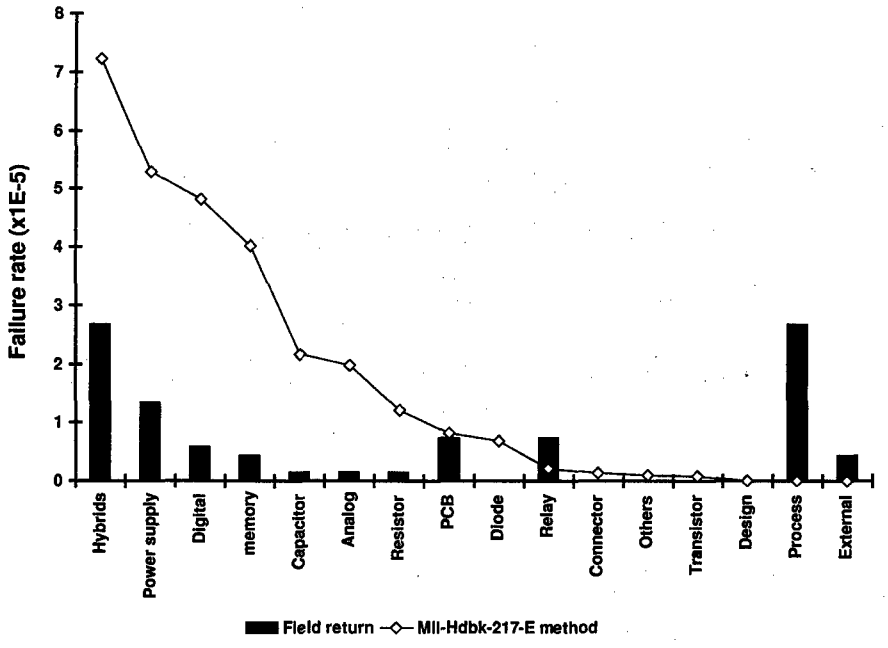



Fig. 1. Example of comparison between field and predictive failure rates of an embedded avionics system

performances.

A previous paper [1] described the proposed methodology based on the reliability tests performed by the parts manufacturers. The data obtained from these standard tests may be useful to estimate the reliability of the COTS used in our applications. Two assumptions are done : all of the intrinsic failure mechanisms are revealed by these tests, and the observed failure mechanisms are the same in the field. The chosen tests are High Temperature Operating Life (HTOL), Temperature Humidity Bias (THB), and Temperature Cycle (TC) or tests similar to these one's. To analyze the results of these reliability stress tests, probabilistic and physical laws are used, like Arrhenius model for temperature effects, Coffin-Manson model for the thermal fatigue effects, Gunn or Peck models for the influence of humidity rate and Poisson or $\chi^2$ distribution laws.

We have successfully studied its feasibility on two electronic components. We have confirmed too the inadequacy between field return data and classical reliability prediction methods and the accordance between our deterministic approach and the field return data. Furthermore, some critical points have been pointed out, as the inaccuracy of the use environment knowledge which may lead to a great variation in the predicted reliability data.

## 4. Critical analysis

In the feasibility study, we chose two plastic encapsulated parts, respectively in small outline (SO8) and lead chip carrier (PLCC84) packages, in order to take into account and assess the humidity rate influence on reliability. When we analyze the obtained results, we may see that the humidity induced failure rate accounts for only an infinitesimal percentage of the whole failure rate. This percentage is independent of the chosen model [5]. According to our field return experience, we may neglect the humidity influence in the calculations.

We use the modified Coffin-Manson relation [6] to assess the thermal fatigue effects. The thermal cycling, and then the thermal fatigue, seems to be the most influential phenomenon on the parts reliability.

We can write the failure rate as :

$$\lambda = \lambda_T + \lambda_{CF} + \lambda_{RH} \quad (1)$$

where $\lambda_T$, $\lambda_{CF}$ and $\lambda_{RH}$ are the failure rates induced by temperature, thermal cycling and humidity rate respectively.

According to the above comments, Eq. 1 becomes :

$$\lambda \approx \lambda_T + \lambda_{CF} \quad (2)$$

If we consider a daily $\Delta T$ of 30°C, we obtain with an activation energy $E_a$ of 0.7 eV :

$$\lambda_T \approx 0.6\ \lambda \text{ and } \lambda_{CF} \approx 0.4\ \lambda \quad (3)$$

On the contrary, with a daily $\Delta T$ of 80°C, the ratio becomes :

$$\lambda_T \approx 0.25\ \lambda \text{ and } \lambda_{CF} \approx 0.75\ \lambda \quad (4)$$

Accurate knowledge of operational daily $\Delta T$ may greatly influence the FIT (failure in time) predictive values. This is the major key point of the proposed methodology.

To confirm this assumption, table 1 shows the influence of the operational temperature gradient on the FIT of the components chosen in the preliminary study (component A in SO8 package, components B1 and B2 in PLCC84 package and different mission profiles [1]).

Table 1 : Coffin-Manson calculated FIT vs operational daily $\Delta T$ values

| **Gradient $\Delta T$** | **30** | **40** | **60** | **80** |
|---|---|---|---|---|
| Component A | 0.9 | 1.6 | 3.6 | 6.4 |
| Component B1 | 12.6 | 22.5 | 50.5 | 90 |
| Component B2 | 5 | 9 | 20 | 36 |

In order to solve this point, the implementation of Time Stress Measurement Devices (TSMD) into the equipment racks may be useful.

Calculating $\lambda_T$ must be performed as accurately as possible too. According to the Arrhenius law, we have to know accurately the mean junction temperature of each part on an electronic board (a 10°C increase induces a twice higher $\lambda_T$). However, some difficulties appear as :

- the major unknown of the Arrhenius law is the activation energy $E_a$, typical of the intrinsic failure mechanism and part technology. Due to the lack of knowledge of accurate $E_a$ for every technology and failure mechanism, we have to use only average activation energies.
- a part can be heated by a neighboring one's and therefore its ambient temperature won't be the ambient temperature inside the box,
- in some specific cases, the equipment packaging will change significantly the component standard thermal paths. Then, it will be impossible to use some parameters given by the part manufacturer (e.g. junction to ambient thermal resistance $\Theta_{JA}$). This problem is encountered inside equipment where the natural convection is negligible, as in land military applications.
- the mean junction temperature of a part depends on the mean dissipated power by the part during its life. This mean power depends on the part mission profile which often differs from the equipment one's.

In order to solve this problem, we have implemented an accurate thermal assessment methodology for the electronic equipment design. This methodology is centered around a thermal simulation tool which allows us to calculate the junction temperature of each part before the realization of the first prototype. This tool takes into account all the parameters necessary to calculate accurately the junction temperatures and their behavior (accurate 3D representation of the equipment, thermal convection, thermal conduction, thermal radiation, thermal inertia of the materials, external environmental conditions, mission profile). The results of this simulation may be confirmed when necessary on the first prototype measuring either the case temperature or the junction temperature of a particular part. These results (mean junction temperature and junction temperature gradients) are used directly as input parameters of the used physical laws.

## 5. Application on an avionics CPU unit

We choose to apply our method to an avionics CPU unit in order to compare the obtained results with other classical methods (MIL-HDBK-217, CNET RDF93) used to assess its reliability and field reliability results. For the chosen CPU unit (made with two cards called here C1 and C2), the ratio between Mil-Hdbk-217 failure rate and field failure rate is over 5. We also use a modified CNET

RDF93, called CRDF, using corrective factors, for which calculated failure rates are today closed to operational failure rates.

In 1996, our field return database registered 16 component failures for the CPU unit during 1.4 $10^6$ fly hours. Our reliability assessment method having been applied on active components only, we registered 8 component failures in this family, 4 for each card. Table 2 shows the obtained operational failure rates, with 60% confidence level. We define $\lambda_G$ as the global failure rate and $\lambda_A$ as the active components failure rate.

Table 3 presents the obtained results with the classical methods for the CPU unit.

Table 2 : Field failure rates for the considered CPU unit ($10^{-6}$ $h^{-1}$)

| | $\lambda_G$ | $\lambda_A$ |
|---|---|---|
| Card C1 | $4.79 < \lambda < 9.28$ | $1.7 < \lambda < 4.97$ |
| Card C2 | $3.51 < \lambda < 7.61$ | $1.7 < \lambda < 4.97$ |
| CPU unit | $11.55 < \lambda < 18.76$ | $4.16 < \lambda < 8.46$ |

In order to apply our method, we defined an average operational environment, described in table 4. Moreover, to take into account operational hours and not only fly hours, we have to apply a 1.5 factor to the predictive failure rates.

In these conditions, we obtained a failure rate $\lambda_A$ of $3.5.10^{-6}$ $h^{-1}$.

Using a daily $\Delta T$ of 80°C, a failure rate $\lambda_A$ of $6.9.10^{-6}$ $h^{-1}$is calculated, twice the previous one's, but in the same order of magnitude.

Table 3 : Predictive failure rates with classical methods ($10^{-6}$ $h^{-1}$)

| | $\lambda_G$ | $\lambda_A$ |
|---|---|---|
| CRDF | 19.65 | 4.34 |
| Mil-Hbdk-217 | 49.1 [a] | 48.9 [b] |

[a] Mil-Hdbk-217 E notice 1 with modified $\pi_Q$
[b] Mil-Hdbk-217 F notice 1

Table 4 : Average mission profile

| | |
|---|---|
| Ambient temperature | 45°C |
| Humidity rate | 50% |
| Daily $\Delta T$ | 1 cycle of 30°C |

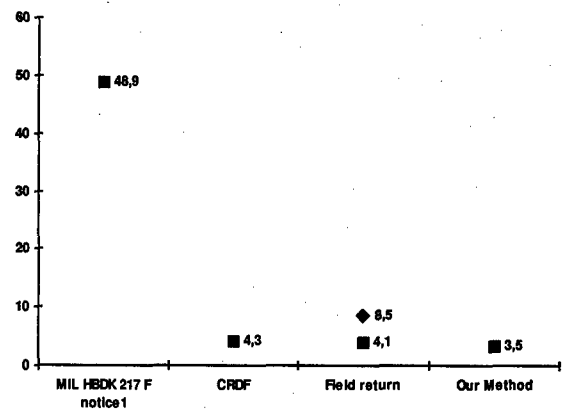



Fig. 2. Failure rates $\lambda_A$ obtained with various methods ($10^{-6}$ $h^{-1}$)

The results are summarized in figure 2.

*The predictive calculated FIT with the presented methodology is similar to the field one's.* The proposed methodology is then valid to assess electronic parts reliability.

We use it to discriminate two potential electrically equivalent parts, choosing the part manufacturer having either the best reliability approach or the best reliability results.

## 6. Conclusions

The estimated failure rate of the avionics CPU unit studied is similar to the field failure rate. The input data of the proposed method being parts manufacturer data, discrimination between two electrically equivalent sources may be easily achieved. However, we have highlighted some critical points such as the use of average activation energies and the accurate knowledge of the profile mission. In order to solve this particular point, the implementation of Time Stress Measurement Devices (TSMD) into the equipment racks may be useful. The reliability assessment methodology will be implemented on future electronic equipment design.

## Acknowledgments

We want to acknowledge Mr Arnaud Meneut for field data analysis and calculations.

## References

[1] Charpenel P. and al., "An other way to assess electronics part reliability", Microelectron. Reliab., Vol.37, n°10/11, Proc. ESREF 1997
[2] Mil-Hdbk-217 F
[3] CNET RDF 1993
[4] Cushing M.J. and al., "Comparison of electronics-reliability assessment approaches", IEEE Trans. Reliability, Vol.42, n°4, 1993
[5] Internal report
[6] Lau J.H., "Thermal fatigue life prediction of flip chip solder joints by fracture mechanics method", Eng. Fracture Mechanics, Vol.45, n°5, pp. 643-654, 1993.

PERGAMON

Microelectronics Reliability 38 (1998) 1177–1182

MICROELECTRONICS RELIABILITY

# Improved reliability of bistable circuits by selective hot-carrier stress reduction

A.G.M. Das[a], S. Johnson[b]

[a] Medical University of Southern Africa, MEDUNSA, 0204, South Africa.
[b] School of Engineering, University of Durham, DH1 3LE, UK.

*Abstract:*

The performance and reliability of submicron CMOS circuits have been affected by hot-carrier stress induced degradation. Three common forms of CMOS latch circuits designed using a 0.7-micron commercial process have been considered in a comparative study of the stress levels experienced by individual devices in the circuit. Average stress levels on all the devices over a typical simulation cycle was used to assess the life-time and the reliability of the circuits. We describe a technique that was used to identify the devices having higher than normal stress which may consequently degrade at a faster rate than other devices leading to an early failure of the circuit. We have developed design techniques that can be used to reduce the stress levels in identified devices. The improvements in life-time and reliability have been assessed and analysed. The best circuit configuration to reduce hot-carrier stress induced degradation has been identified.

© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Hot-carriers have been known to degrade the performance of submicron CMOS bistable circuits when operated under dynamic conditions [1]. The mechanisms of hot-carrier degradation and the stress conditions that cause this degradation are well understood [2]. The consequences of device parameter shifts on performance and reliability of the circuits are however hard to predict, as the stress experienced by each transistor in a circuit depends on the circuit configuration and operating condition. The stress is also a dynamic phenomenon and is determined by factors such as operating speed.

In this paper we describe a technique which we have used to evaluate the relative and absolute stress levels experienced by all transistors in a number of bistable circuits. The stress on each device is translated into parameter shift in the devices. Using substrate current ($I_{sub}$) as a monitor, the hot-carrier stress experienced by each device in the bistable circuits has been modelled. Average stress level on all the devices in a circuit was used to evaluate the lifetime and the reliability of latch circuits. A number of techniques can be employed to reduce the stress on certain critical devices and thereby improve the overall reliability of the circuit.

## 2. Circuit level stress simulations

Three common forms of CMOS latch circuits designed using a 0.7-micron commercial process have been selected for circuit level dynamic stress simulations and are given in figure 1. Quasi-static analysis has been used with typical dynamic simulation cycles of 20 ns and the effective stress period for each device was evaluated as 5 ns, which is 25% of the simulation period. The average stress factor $S_{ac} = \Sigma I_{sub}/W$ was evaluated for a simulation period where the width $W$ for all the devices were 1 micron. The average stress factors (averaged for all devices in the circuit) calculated for the three latch circuits were 7.38 A/m, 4.80 A/m and 6.59 A/m respectively.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00084-5

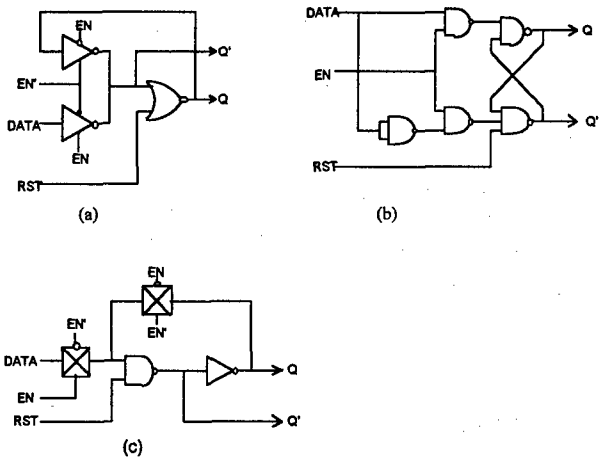



Figure 1. The three bistable circuits selected for stress simulations

It was established from static degradation characterisation simulations and by comparing results of stress induced degradation experiments by Lunonburg [3] that a 20% degradation in device parameter is equivalent to a static stress period of 1110 hours under peak stress conditions. Under dynamic stress conditions, this translates into 4440 hours of AC stress because the devices in a latch will be under stress for 25% of the dynamic stress simulation period and will produce 14% degradation in drain current.

### 3. Selective stress reduction techniques

A transistor level circuit for the first D-type latch DL1 is shown in figure 2. The highly stressed devices were identified as N5 and N6 having a stress level of 21.1 A/m each. Device N3 also has a higher than normal stress (9.9 A/m) and N1 has a normal stress level of 5.4 A/m. Devices N2 and N4 have negligibly small stress levels. It can be seen from the circuit diagram for DL1 that these transistors having high levels of stress are at the output nodes of the latch.

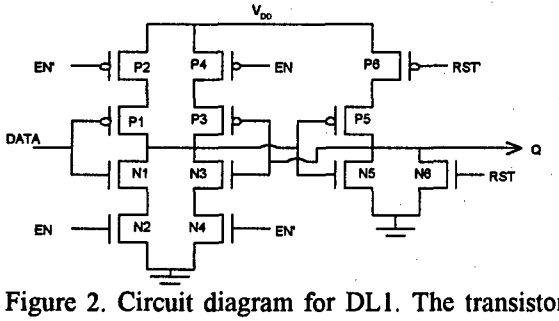



Figure 2. Circuit diagram for DL1. The transistors N3, N5 and N6 have higher levels of stress.

Many researchers [3-5] have shown that hot carriers are generated close to the drain of the nMOS device as the channel electric field has a pronounced peak in this region. While the drain voltage is *high* and the gate voltage is ramped, hot electrons are generated in the channel due to impact ionisation. By reducing the peak of the channel electric field and moving the peak away from the drain, a reduction in the stress level and the hot carrier generation can be achieved. The devices N2 and N4 have very much reduced stress levels even though the drain currents are the same as that for N1 and N3 respectively since N1 is in series with N2 and N3 is in series with N4. The devices N2 and N4 are at the lower ends of the ladders and the *ON* resistance for N1 and N3 act as load at the drain end of these devices which has the effect of reducing the stress on these devices. Similar observations were also made for the other two latch circuits. The *ON* resistance of the upper device acts as a load spreading the electric field and hence diverting the peak away from the drain of the device for which the gate voltage is ramped. Since the stress level on N2 is very much lower than the normal stress, it is possible to reduce the stress on N1 simply by interchanging N1 with N2. A similar technique was used to reduce the stress on N3 by swapping N4 with N3 such that the *ON* resistance of the upper device acts as a load for the device having higher stress level there by reducing the stress level.

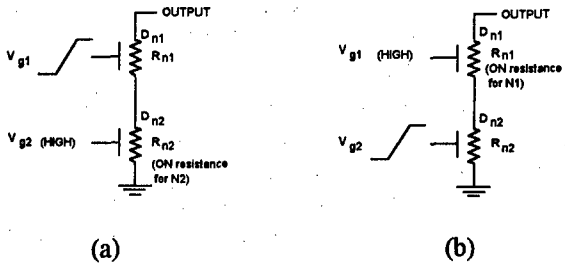



Figure 3. The channel resistance for two of the nMos devices used in the latch circuit (a) the series connected nMOS transistors N1 and N2 before swapping and (b) after swapping.

The nMOS devices N1 and N2 are connected in series for this latch as shown in figure 3(a) before the transistors are swapped. When $V_{g2}$ is *high* the channel resistance $R_{N2}$ is the *ON* resistance of $N_2$ and if $V_{g1}$ is *low*, device N1 is *OFF* and has a very large channel resistance; $V_{OUT}$ is *high* and $V_{DN2}$ is

approximately at ground potential and the full output voltage is dropped across $R_{N1}$. When $V_{g1}$ is now ramped *high*, the peak of the field is at the drain end of N1 and will generate drain current and substrate current at high stress levels during a typical operation of the transparent D-type latch.

By exchanging $V_{g1}$ and $V_{g2}$, we have the *ON* resistance $R_{N1}$ as load to $R_{N2}$ as shown in figure 3(b). When $V_{g2}$ is ramped while $V_{g1}$ is *high*, the *ON* resistance $R_{N1}$ acts as a load for N2 which reduces the peak of the electric field near the drain of N2 and hence reduces the stress and the substrate current for N2. It is also possible to reduce the stress on N3 in the second tri-state device by exchanging transistors N3 and N4. The circuit simulations have shown that this is an effective method for reducing the stress.

Devices N5 and N6 are in parallel and it is not possible to reduce the stress on these devices using the technique described above as the output voltage is appearing across the drain and source of these devices. When we use the minimum dimensions allowed by the technology design rules, we have a very narrow channel, 1-micron width, for the transistors between the drain and the source making the current density and the electric field magnitudes very high. When the channel length of the devices N5 and N6 were increased to 1.5 micron to reduce the field and the widths doubled, the impact ionisation per unit area of cross-section of the channel was reduced. The increase in widths have given rise to an increase in peak saturation currents for the devices and the substrate currents were also increased. But the substrate current generation was not increased proportionately as the electron temperature is shown to be lower in a wider device [4] and hence the normalised substrate current generated also was reduced.

The channel lengths of the devices N5 and N6 were increased to 1.5 μm and the circuit was simulated to measure the stress over a cycle. The increase in channel length of the two devices may not affect the overall response of the latch adversely as we have already doubled the channel width for these devices while all other devices in the circuit have normal widths. The reduction in the stress levels for the devices have been determined using circuit simulations and was found that the stress levels have been reduced to 11.5 A/m from 16.8 A/m due to the increase in length. The stress factors for the nMOS devices in DL1 before and after the design modifications are shown as a bar graph in figure 4.

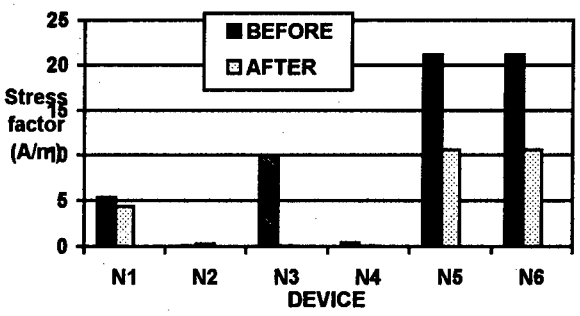



Figure 4. Stress factors for devices in DL1 before and after design improvements

The life-time for each of the transistors was calculated using the life-time model developed for AC stress simulations [6] and is given by,

$$\tau = H(\Delta D / D)_f W (\Sigma I_{SUB})^{-m} (\Sigma I_{DS})^{m-1} \quad \text{---(1)}$$

where the degradation $(\Delta D/D)_f$ = 0.1 or 10% for failure, the drain current $(\Sigma I_{DS})$ and the substrate current $(\Sigma I_{SUB})$ are those for one simulation cycle; $H$ and $m$ are constants. The stress factors and life-times for the nMOS devices in the latch before and after design improvements are tabulated in table 1. The results in table1 show that many devices in this circuit has very high stress level contributing to reduced life for the devices and hence the circuit is unreliable. Even though we have achieved a stress reduction of about 50% for N5 and N6, the stress levels remain unacceptably high leading to an early failure of these devices.

Table 1. Stress factors and life-time for devices in DL1 before and after design improvements

| | Stress factor (A/m) | | Life-time (Hours) | |
|---|---|---|---|---|
| Device | Before | After | Before | After |
| N1 | 5.4 | 4.4 | $1.7 \times 10^5$ | $3.9 \times 10^5$ |
| N2 | 0.1 | 0.3 | $1.0 \times 10^{10}$ | $6.4 \times 10^8$ |
| N3 | 9.9 | 0.1 | $1.2 \times 10^3$ | $9.6 \times 10^8$ |
| N4 | 0.4 | 0.1 | $1.4 \times 10^7$ | $9.8 \times 10^8$ |
| N5 | 21.1 | 10.6 | $2.1 \times 10^3$ | $5.5 \times 10^3$ |
| N6 | 21.1 | 10.6 | $2.1 \times 10^3$ | $5.5 \times 10^3$ |

Selective stress reduction techniques were also employed for two other bistable circuits and the results are shown in the bar graphs in figures 5 & 6.

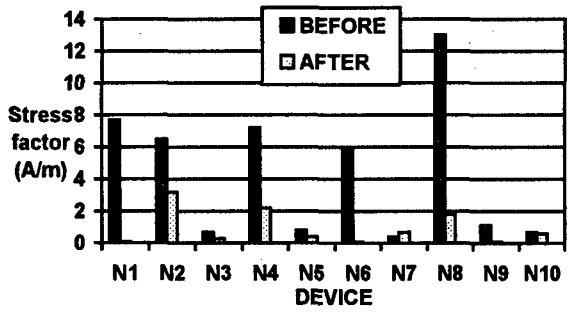



Figure 5. Stress factors for DL2 before and after design changes

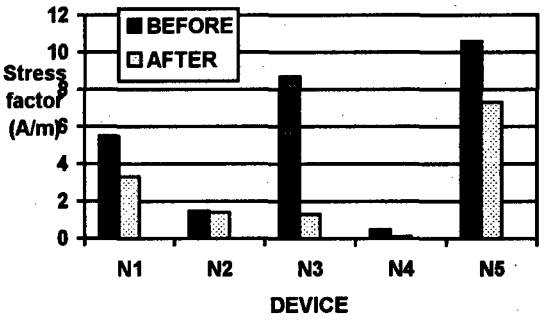



Figure 6. Stress factors for DL3 before and after design changes

It can be seen from figure 5 that the transistor N8 had the highest stress level (13 A/m) before design improvement. Since this device is at the output node of the latch circuit, it will suffer maximum degradation and can lead to circuit failure. With the design improvements, we have reduced the stress on this device to 1.8 A/m and the life-time also has been improved by a factor of 100. The overall improvement in the life-time of the circuit will be evaluated in the next section.

Figure 6 shows the simulation results for DL3. The highest stress experienced by the transistor N5 is at the output node of the latch. Our design improvements were aimed at improving the stress on this device as well as that on N3. We have reduced the stress on N3 to 1.3 A/m and the stress on N5 has been lowered to 7.3 A/m. We have not achieved the level of improvement in the life-time for this device by reducing the stress and so this device is the critical one which might lead to failure of the circuit.

Based on the stress levels on the devices, the device parameters were degraded and the circuit re-simulated to determine the improvements in the stress levels. Quantitative analysis of the improvements in lifetime and reliability of the circuits are discussed in the next section.

### *3.1 Estimating the improvements in life-time*

The average stress factors and life-times for the circuits are tabulated in table 2. When a circuit is operated normally, the end of life may be due to a critical, highly-stressed device for which the timing parameters have been degraded due to hot carrier effects. The stress level and the relative position of the device in the circuit can be critical for such circuits.

If the device under stress is at the output node of the latch circuit, the *rise* and *fall* times and the propagation delays are affected adversely; if the device is at the *data* input path, the propagation delays are affected. The stress on the identified critical devices and the corresponding life-times for the circuits are given in table 3.

Table 2. Stress factors and lifetime for the latch circuits before and after design modifications

| | Average Stress factor (A/m) | | Life-time (Hours) | |
|---|---|---|---|---|
| Circuit | Before | After | Before | After |
| DL1 | 9.7 | 4.4 | $1.4 \times 10^4$ | $1.7 \times 10^5$ |
| DL2 | 4.4 | 0.9 | $1.5 \times 10^5$ | $3.8 \times 10^7$ |
| DL3 | 5.4 | 2.7 | $5.9 \times 10^4$ | $8.4 \times 10^5$ |

Table 3. Maximum Stress factor on the critical device and life-time for the latch circuits before and after design modifications

| | | Max. Stress factor (A/m) | | Life-time (Hours) | |
|---|---|---|---|---|---|
| Circuit | Device | Before | After | Before | After |
| DL1 | N5 | 21.1 | 10.6 | $2.1 \times 10^3$ | $5.5 \times 10^3$ |
| DL2 | N8 | 13 | 1.8 | $1.4 \times 10^4$ | $1.1 \times 10^6$ |
| DL3 | N5 | 10.6 | 7.3 | $5.5 \times 10^3$ | $1.8 \times 10^4$ |

A comparison of the life-times shown in tables 2 and 3 for the latch circuits reveal that the life-time evaluated using the stress on the critical device in a latch is much shorter than that calculated using the averaged stress value. This result shows that the eventual failure of the circuit may occur relatively earlier if the critical device in the circuit degrades faster than the average degradation rate. Hence it is very important to assess the stress levels

on identified critical device and to address special techniques to reduce the stress for this device.

## 4. Analysis of the reliability improvements

The end of an operational life of a circuit may be due to the failure of the circuit to meet the characteristic timing parameters of the circuit as a result of the degradation of one or more of the highly stressed critical device. A device is considered critical if it is in the signal path of the circuit which can give rise to increased propagation delays and transition times. It was shown that the *most critical device* in a bistable could be found at the output node of the latch circuit. There are two main reasons for this: when the bistable is operated under dynamic conditions, the gate of the output device experiences a constantly changing voltage and the drain-to-source voltage of this device is also changing, giving rise to the condition necessary for peak substrate current generation [1]. It was shown that the rise and fall times of a bistable circuit were affected to a maximum degree when the device at the output degrades faster than other devices in the circuit. Hence the circuit life-time for the three latch circuits were also evaluated using the stress factor on the most critical device in that circuit.

### *4.1. Reduction in stress factor*

The improvement in the life-time and the reliability of the bistables were assessed by evaluating the reduction in average stress levels as a result of design improvements. The average stress factors for the three latch circuits before design improvements are 9.7 A/m, 4.4 A/m and 5.4 A/m respectively. The relatively higher stress level on DL1 has resulted in a higher rate of degradation and hence a shorter life-time for this circuit. The second latch circuit, DL2, has the lowest stress level compared to the other two circuits. Further improvement in the stress levels were achieved by design improvements and the stress factor has been lowered to 0.9 A/m. The stress level for the third latch circuit was also lowered to an acceptable value (2.7 A/m) and the reduction amounts to 55%, 80% and 50% respectively for the three circuits.

The failure of a latch may be due to a critical device having higher stress degrading at a faster rate than others. Hence we have identified the critical device in the latch circuits and have evaluated the time-to-failure of the critical device leading to circuit failure.

Reduction in stress factor for DL1 (critical device: N5) = 50%
Reduction in stress factor for DL2 (critical device: N8) = 86%
Reduction in stress factor for DL3 (critical device: N5) = 31%
We have achieved significant improvement in the stress factors for the critical devices in the three circuits. These improvements will be reflected as life-time and reliability improvements for the circuits and are discussed next.

### *4.2 Improvements in life-time*

The improvements in the life-times for the latch circuits after the design modifications were found to be 1100%, 2500% and 1300% respectively for the three circuits. We have also evaluated the life-times of the circuits using the stress experienced by the critically stressed device as given in table 3. The improvement in life-times for the three circuits were now found to be 162 %, 7800% and 227% respectively.

### *4.3 Reliability improvements*

The reliability of a circuit having constant failure rate has been defined as,

$$R(t) = \exp(-t/\tau) \quad \text{--- (2)}$$

where $\tau$ is the life-time of the circuit. The reliability of the circuits was evaluated using the life-times determined in the previous section. The normal time-to-fail was defined on the basis of a deterioration in drain current by 10% and for the latch circuits, this was determined as 3000 hours of operation at maximum stress. The reliability of each of the latch circuits was evaluated based on the failure of the critically stressed device in each circuit tabulated in table 3. Reliability evaluated at the end of 3000 hours of operation for the three latch circuits before and after the design improvements to reduce the hot-carrier stress are listed below:
Reliability for DL1 (before modification) [t=3000 hours] = 24%
Reliability for DL1 (after modification) [t=3000 hours] = 58%. The life-time of the device was only 2100 hours before the design modifications and so the circuit would have failed before 3000 hours. With the design improvements to reduce the hot-carrier stress, we have achieved a reliability improvement of 34% for this circuit at the end of

3000 hours of operation. The reliabilities for the other two circuits were 81% and 58% before and 99% and 85% after the design modifications respectively.

The reliability of the circuits was also evaluated based on the average stress factor (averaged for all devices in the circuit) and the corresponding life-times are shown in table 4. The average life-times for the circuits were found to be close to 5 years and so the reliability was evaluated at t = 5 years for all the circuits and are tabulated in table 4.

Table 4. Reliability for the three circuits before and after design improvements

| Circuit | Reliability (t = 5 years) (%) | |
|---|---|---|
| | Before modification | After modification |
| DL1 | 4 | 77 |
| DL2 | 75 | 99 |
| DL3 | 47 | 95 |

The reliability of DL1 is lowest (4%) and it would have failed before the 5 year period. After the design modifications, the reliability has improved to 77%. The reliability for DL2 and DL3 were much higher even before the design modifications; it was further improved by 24% and 48% respectively by design modifications to reduce hot-carrier stress.

## 5. Summary and conclusions

We have assessed the reduction in stress levels for the selected bistables as a result of the design improvements in order to reduce hot-carrier degradations. The improvement in life-time was used to assess the reliability improvement for the latch circuits. The average stress levels on each of the selected three bistable circuits had a wide range of values during a typical operating cycle. The first bistable DL1 had the highest stress level and the design improvements have been shown to reduce the stress and improve the life-time for this circuit more than the other two. The reliability improvement for this circuit is better than the others, an improvement of 73% over a five year period. The second bistable had the minimum stress and hence best reliability before improvements and have achieved further improvements. This circuit configuration has been identified as the best for hot-carrier reliability. An improvement in reliability for the third bistable circuit also was possible using selective stress reduction techniques.

Our investigation into the reliability and life-time of the bistable circuits have revealed that the stress levels and the corresponding hot-carrier degradation of the devices in a circuit depend on the position of the device relative to other devices and the sequence at which the devices are operated. The techniques developed to reduce the stress levels in identified devices have improved the reliability and the life-time of the circuit. We have also identified the best circuit configuration to reduce hot-carrier degradation. For high reliability VLSI circuits, one must use bistable circuits that have been identified as having best performance and improved reliability in spite of hot-carrier degradation effects.

## Acknowledgements

I acknowledge financial assistance given to me by the Medical University of Southern Africa for the research work and for attending ESREF 98 at Copenhagen. Financial assistance given by De Beers Chairman's Fund Educational Trust is also acknowledged.

## References

[1] Leblebici Y., *Modeling and Simulation of Hot-Carrier Induced Device Degradation in MOS Circuits,* IEEE J SS circuits, v 28, No: 5, May 1993, pp 585-595.

[2] Hu C., Tam S., Hsu F, Ko P., Chen T. and Terrill K., *Hot electron induced MOSFET degradation- model, monitor and improvements,* IEEE Tran. ED-32, no 2, Feb 85.

[3] Lunenborg M.M., *MOSFET Hot-Carrier Degradation, Failure mechanisms and models for reliability circuit simulation,* Ph.D. thesis, University of Twente, Netherlands, 1995.

[4] Ning T.H., Cook P.W., Dennard R.H., Osburn C.M., Schuster S.E. and Yu H., *1 μm MOSFET VLSI Technology: Part IV- Hot-Electron Design Constraints,* IEEE tran ED-26, no 4, April 79, pp 346-353.

[5] Liang C., Gaw H. and Cheng P., "An analytical model for self-limiting behavior of hot-carrier degradation in 0.25-μm n-MOSFETs", IEEE EDL-13, no 11, Nov 92, pp 569-571.

[6] Das A.G.M., *Effect of Wearout Processes on the Critical Timing Parameters and Reliability of CMOS Bistable Circuits,* Ph.D. thesis, University of Durham, UK, 1997.

PERGAMON

Microelectronics Reliability 38 (1998) 1183–1186

MICROELECTRONICS RELIABILITY

# A study of NMOS behavior under ESD stress: simulation and characterization

Albert Z. Wang, Chen Tsay, Amit Lele and Peter Deane

National Semiconductor Corp., 2900 Semiconductor Dr., MS: A2-595, Santa Clara, CA 95052, USA

ABSTRACT

A full-scale simulation-aided ESD design methodology was used to design a group of NMOS ESD protection units. Silicon results match the simulation data quite well. Both simulation and measurement data show good ESD performance uniformity across NMOS poly finger length and finger number in ladder structures in a large range. Optimal layout pattern for ladder structures was obtained with the aid of simulation.

© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

NMOS structures have been widely used to protect integrated circuit (IC) chips against electrostatic discharge (ESD) damage [1]. A great deal of work has been done focusing on improvement and rules of the thumb of NMOS ESD protection designs. However, inconsistency often exists in the publications due to lack of a mature theory. In contrast to the IC designs where computer-aided design (CAD) plays a major role, simulation has not been really used in the practical designs of ESD structures because rare convincing results have been observed. Therefore, the experience-oriented trial-and-error method still dominates the mainstream ESD design community.

In this talk, we report results of a predictive NMOS ESD design work by use of a home-developed full-scale simulation methodology. The simulation was confirmed by silicon measurements. Data of NMOS ESD performance uniformity on finger length and finger numbers, which do not agree with previous papers, are to be discussed. Issues on optimal layout of NMOS ladder structure will also be presented.

## 2. Simulation

In order to be able to predictably design ESD protection devices, a full-scale simulation method was developed at National Semiconductor Corp. [2], which consists of process, device and circuit simulations. Unlike all previous reported work [3], this methodology, upon full calibration, enables us to conduct ESD protection designs with much confidence by using well-proven, existing models of processing, device, circuit and ESD (including HBM,

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00117-6

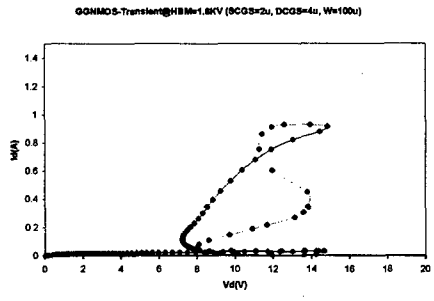



Fig. 1 Transient I-V curve of a GGNMOS.

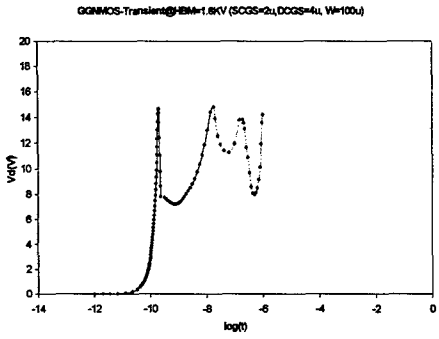



Fig. 2 Semi-log transient I-V curve of a GGNMOS.

MM & CDM) without any assumption. The method was applied to designs of ground-gate (GGNMOS) and gate-coupling (GCNMOS) structures with the goal to investigate such critical characteristics as, I-V curve, triggering, snapback, timing, gate-coupling, and temperature, as well as to choose proper device structures (e.g. drain/source-contact-to-gate-spacing, DCGS & SCGS) and peripheral devices (e.g. capacitor, resistor & diode) etc. With the aid of the complete simulation, single and multiple poly gate finger GGNMOS and GCNMOS structures were optimized for a 0.8um BiCMOS process application. In a GGNMOS example, simulation suggests

Table 1 Data from transient simulation and measurement.

| Devices | GGNMOS | | GCNMOS | |
|---|---|---|---|---|
| Items | Simulation | Silicon | Simulation | Silicon |
| $V_{t1}$ (V) | 14.68 | 12.56 | 7.54 | 6.66 |
| $t_1$ (ns) | 0.2 | -- | 0.42 | -- |
| $V_h$ (V) | 6.92 | 6.48 | 7.41 | 6.08 |
| $V_G^{max}$ (V) | -- | -- | 3.67 | -- |
| $t_G^{max}$ (ns) | -- | -- | 0.32 | -- |

optimized device with SCGS=2um, DCGS=4um and no LDD. Transient simulation results for a GGNMOS with gate finger length W=100um under HBM=1.6KV stressing are shown in Fig.-1 and Fig-2 with the extracted data summarized in Table-1 along with the measured data. Compared to the pulse rise time of about 15 ns in HBM event, the very short triggering time of $t_1$=0.2ns ensured fast turn-on of the GGNMOS and the device passed 1.6KV HBM stressing. Shown in Figs.-3, 4, & 5 are results of a

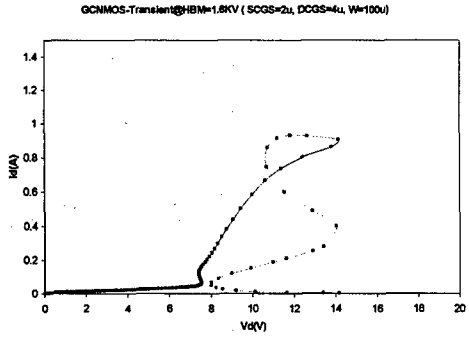



Fig. 3 Transient I-V curve of a GCNMOS.

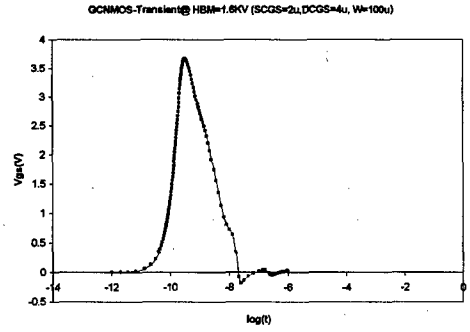



Fig. 4 Transient gate voltage of a GCNMOS.

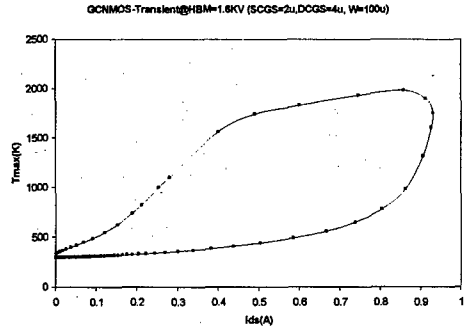



Fig. 5 Max-lattice-temperature *vs.* $I_{ds}$.

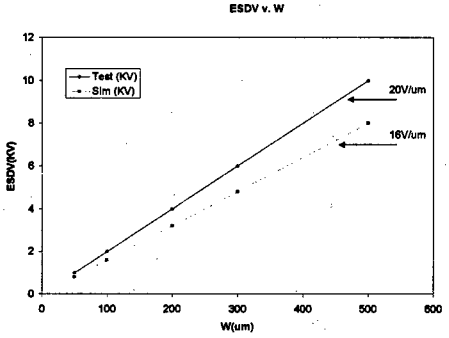



Fig. 6 ESDV level *vs.* finger length.

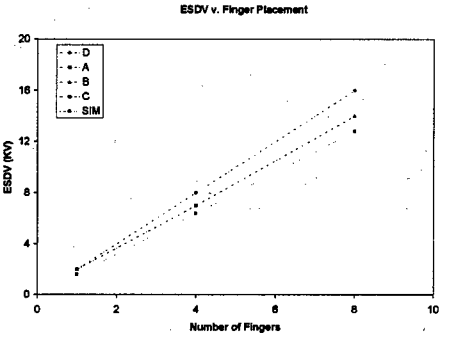



Fig. 8 ESDV level *vs.* layout patterns.

GCNMOS with SCGS=2um, DCGS=4um, W=100um, C=0.1pF and R=10KΩ under HBM=1.6KV stressing. The trigger voltage was designed at $V_{t1}$=7.5V with $t_l$=0.42 ns, longer than the GGNMOS case due to the coupling effect. In order to achieve efficient coupling, the maximum transient gate bias was properly designed with a fast rising time of t=0.32 ns. The transient gate biasing duration should be set long enough to ensure full triggering, but short enough so that no excess leakage current may continue after the ESD event. The simulation data are to be compared with silicon results in the following section.

## 3. Experiment & discussion

Summarized in Table1 are simulation and silicon data measured using a Tektronics 370 Curve Tracer for a GGNMOS and a GCNMOS, both are one-finger devices. The matching is quite satisfactory

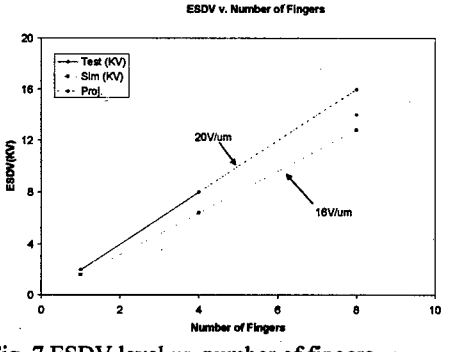



Fig. 7 ESDV level *vs.* number of fingers.

because full calibration was done in advance. The simulated trigger voltage, $V_{t1}$ is higher than the measured data because simulation used transient pulse while the curve tracer produced a quasi-DC stressing. A group of one-finger GGNMOS devices with various poly finger length was designed to investigate the ESD performance (ESDV) *verse* finger length property. The HBM zapping results are plotted in Fig.-6 along with the simulated data. The test data are slightly higher than the simulation ones, indicating further calibration needed. Interestingly, very good ESDV uniformity across a wide range of finger length was observed which argues the common understanding that too long a finger may cause current crowding and lead to premature ESD damage, therefore no finger longer than 150um is recommended [4]. It is believed that the complete simulation contributed to this result. Study on the efficiency of the common ladder structures was also conducted on devices with multiple poly gate fingers. The HBM zapping results are shown in Fig.-7. The data for the 8-finger device (projected 16KV) is incomplete because of the limit of the zapping machine (14KV, passed). However, it strongly indicates that good ESDV uniformity across a large finger number range exists. This is in contrast to the common belief [1, 5] that a ladder structure with more than three fingers usually suffer premature damage because one or more fingers may be damaged before all can be turned on, therefore complicated structures were proposed for high efficiency. Again, authors believe that the optimized structure by simulation is a key in achieving this improvement. The different layouts of ladder

structures were also studied with the HBM zapping results shown in Fig.-8. The structure patterns used are: (A) = BSGDGSGDGSB, (B) = BSGDGSBSGDGSB, (C) = B-DGSGD-B-DGSGD-B, and (D) = BSGD-DGSBSGD-DGSB. The ESDV uniformity across the fingers seems existing for all cases; however, the pattern (D) is superior to the others.

## 4. Conclusion

In summary, GGNMOS and GCNMOS ESD structures were designed with the aid of a full-scale simulation methodology. Silicon measurements match the simulation very well. ESD performance scaling to both finger length and finger number within a large range was found achievable for devices designed by simulation.

## References

[1] A. Amerasekera and C. Duvvury, ESD in Silicon Integrated Circuits, John Wiley & Sons, New York, 1995.

[2] Albert Wang, Internal Technical Report, National Semiconductor Corp., Dec. 1997.

[3] C. Diez, "Modeling and Simulation of Electrical Overstress Failures in Input/Output Protection Devices of Integrated Circuits", Ph.D Thesis, UIUC, 1993.

[4] Private communications with T. Polgreen, ESD Solutions.

[5] T. Polgreen and A. Chatterjee, "Improving the ESD Failure Threshold of Silicided nMOS Output Transistors by Ensuring Uniform Current Flow", Proc. 11th. EOS/ESD Symposium, p167-174, 1989

PERGAMON

Microelectronics Reliability 38 (1998) 1187–1191

MICROELECTRONICS RELIABILITY

# The time of "guessing" your failure time distribution is over!

K. Croes[1], J. V. Manca[1], W. De Ceuninck[1], L. De Schepper[1] and G. Molenberghs[2]

[1]*Limburgs Universitair Centrum, Institute for Materials Research, Materials Physics Division, Universitaire Campus, Wetenschapspark 1, 3590 Diepenbeek, Belgium*

[2]*Limburgs Universitair Centrum, Biostatistics, Universitaire Campus, 3590 Diepenbeek, Belgium*

**Abstract**

Each statistical analysis of reliability data starts with the choice of the underlying distribution of failure times. This choice is of great importance because all conclusions drawn from this analysis will depend on it. Lifetime predictions can vary orders of magnitude depending on the distribution used. Most researchers choose the underlying distribution of failure times rather unfounded: because of "historical" reasons, because everybody uses it,...We developed a method which offers reliability engineers an objective tool for making the distinction between the two most widely used distributions, the lognormal and the Weibull, using a statistical well-founded technique. Essentially, the method comes down to constructing both the lognormal and the Weibull probability plot of the data set under consideration. For each plot, the Pearson's correlation coefficient is calculated. It is shown that the ratio of these two correlation coefficients is a pivotal quantity. Hence, it can serve as a test statistic. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

For many statistical analyses of reliability data, a main concern is whether to fit the data to a lognormal or to a Weibull distribution. The importance of this choice is obvious: the extrapolation to low percentiles (an $x$ %-percentile is defined as the time that $x$ % of the total population of components have failed) is particularly sensitive to the choice of the underlying distribution of failure times.

The choice of the underlying distribution of failure times occurs mostly unfounded: because of "historical" reasons, because everybody works with it,... In the reliability literature one can see that researchers use both distributions to represent failure data of the same failure mechanism. For Time Dependent Dielectric Breakdown (TDDB) experiments, for instance, one often assumes a Weibull distribution, since this distribution fits with the weakest-link character of the breakdown process [1]. Bain [2], however, used the lognormal distribution. Another example deals with electromigration. Here, one often assumes a lognormal distribution. A number of papers have been published to support this assumption. In one of these papers, it is shown that a normal temperature distribution within a sample set produces a lognormal failure rate if the Arrhenius model is assumed [3]. On the other hand, it is shown that a barrier layer can change the distribution of the interconnect electromigration failures [4] and that the length of a test stripe can influence the underlying distribution of electromigration failures

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00083-3

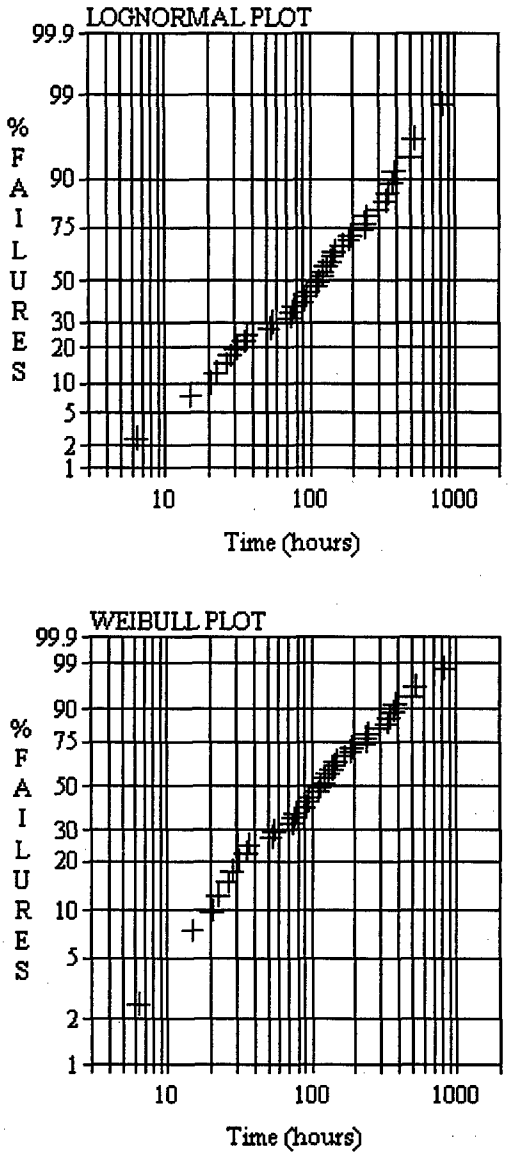



Fig. 1. Lognormal and Weibull probability plot of 40 simulated failure times

[5]. In summary, one can state that some researchers use the lognormal distribution while others use the Weibull distribution in the absence of a theoretical justification.

Figure 1 shows both the lognormal and the Weibull probability plot of 40 simulated failure times. What do you think is the correct failure distribution? You'll have to be a little patient since the underlying distribution function of these failure times will only be revealed at the end of this paper. The method proposed in this paper will make it easy to make this choice.

Before introducing the new method, it should be noticed that it is not the authors' intention to overthrow established physical theories concerning the distribution of failure times. The authors' aim is to provide a statistical tool to distinguish between the lognormal and the Weibull distribution for experiments with unknown failure distribution or unknown failure mechanism

## 2. Dangers of choosing the wrong distribution

The danger of choosing the wrong distribution function can best be seen while comparing the results of a lognormal and a Weibull fit. The estimated 0.01%-percentile for the lognormal fit performed on the failure times of figure 1 was 1.41 seconds, while this estimate amounted to 0.0218 seconds for a Weibull fit. This is a difference of a factor 65!

## 3. Statistical quantity of importance: $\rho_W/\rho_L$

Before introducing the statistical quantity of importance, few definitions have to be made. First, define $n$ as the total number of components in the experiment. Second, $r$ is defined as the total number of failures in that experiment.

The last definition deals with Pearson's correlation coefficient, $\rho$, of two vectors $x$ and $y$ of length $m$. $\rho$ is defined as

$$\rho = \frac{\sum_{i=1}^{m}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{m}(x_i - \bar{x})^2 \sum_{i=1}^{m}(y_i - \bar{y})^2}} \quad (1)$$

with $\bar{x}$ and $\bar{y}$, the respective means of $x$ and $y$. $\rho_W$ is defined as this correlation coefficient of the points on the Weibull plot of the experiment under consideration. $\rho_L$ is defined in the same way but now for the lognormal plot (it should be noted that the above mentioned vector $x$ is the same for both plots, but the vector $y$ is different due to a different scaling

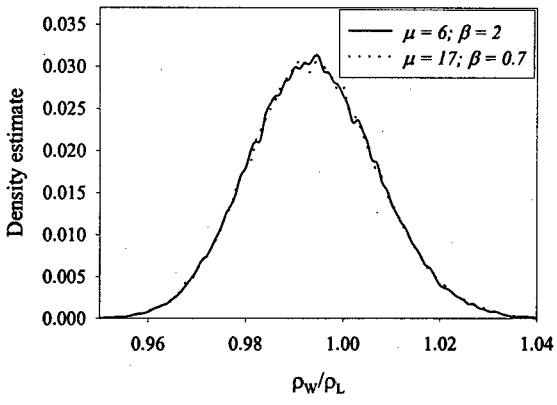



Fig. 2. Two plots of the estimated density functions of $\rho_W/\rho_L$ with $n = 60$ and $r = 40$. The underlying distribution is lognormal. The difference between the two curves is the choice of the values for the scale parameter $\mu$ and the shape parameter $\beta$.

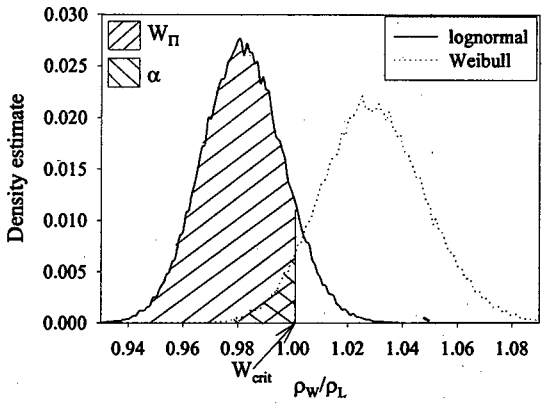



Fig. 3. Two plots of the estimated density functions of $\rho_W/\rho_L$ for data sets with $n = 60$ and $r = 40$. The difference between the two curves is the underlying failure distribution function.

of the y-axis). The statistical quantity of importance is $\rho_W/\rho_L$.

It will be shown that the distribution of the ratio $\rho_W/\rho_L$ over different experiments only depends on $n$, $r$ and on the type of the underlying failure distribution, but NOT on the scale and the shape parameter of the underlying failure time distribution. Thus, it is a pivotal quantity.

Figure 2 supports this statement. This figure shows the result of calculations performed on 20,000 simulated data sets with $n = 60$ and $r = 40$. The underlying failure distribution of all these data sets was chosen to be lognormal. The first 10,000 data sets were simulated with scale parameter $\mu = 6$ and shape parameter $\beta = 2$, while for the other 10,000 data sets, $\mu = 17$ and $\beta = 0.7$ was chosen. For each data set, the ratio $\rho_W/\rho_L$ was calculated. The full line on figure 2 is the estimated density function of these $\rho_W/\rho_L$ for the first choice of $\mu$ and $\beta$, while the dotted line is the estimated density function of these $\rho_W/\rho_L$ for the second choice of $\mu$ and $\beta$. From the figure, it is easy to observe that the estimated density function of $\rho_W/\rho_L$ does not depend on these $\mu$ and $\beta$. A Kolmogorov-Smirnov goodness-of-fit test is performed in order to verify whether there is a difference between the two density functions. The $p$-value of this test was > 0.05, which supports the above mentioned statement. More plots were inspected for more combinations of $n$ and $r$, of the underlying failure distribution (Weibull or lognormal) and of $\mu$ and $\beta$. For each plot, a Kolmogorov-Smirnov test was performed. All tests resulted in the same conclusion: $\rho_W/\rho_L$ is a pivotal quantity.

## 4. Method in theory

Figure 3 shows the estimated density functions of the calculated $\rho_W/\rho_L$ values of 20,000 simulated data sets, 10,000 for each curve, with $n = 60$ and $r = 40$. The difference between the two curves is in the choice of the underlying failure distribution (from section 3, it is known that the choice of the scale and the shape parameter is of no importance). The full line corresponds to the simulations with an underlying lognormal distribution, while the dotted line represents the simulations with an underlying Weibull distribution. It is clear that $\rho_W/\rho_L$ tends to be higher if the underlying failure distribution is Weibull. This is to be expected: Weibull samples give a higher correlation for the Weibull plot where lognormal samples give the opposite. Intuitively, testing Weibull versus lognormal can be done as

follows. First, calculate $\rho_W$ and $\rho_L$ and determine the ratio $\rho_W/\rho_L$. If $\rho_W > \rho_L$, then choose Weibull and if $\rho_W < \rho_L$, then choose lognormal. The following technical section shows how to perform the test in a statistically efficient way. Note that the method proposed here is incorporated in a commercially available software package[6].

## 5. Technical details of the test

Testing the null hypothesis $H_0$ : Weibull versus the alternative $H_A$ : Lognormal can most easily be done using the technique depicted in figure 3. The next 4 steps should be followed:

1) Choose a significance level $\alpha$ (the significance level of a test is defined as the probability that $H_0$ is rejected given that $H_0$ is true). A significance level should be chosen low.
2) Find the critical value $W_{crit}$ of the test. This is the value on the x-axis which divides the estimated density curve of $\rho_W/\rho_L$ for Weibull samples into two parts: the left part containing $\alpha$*100% of the total area under the curve.
3) For a given data set, the hypothesis $H_0$ : Weibull is rejected when the observed ratio $\rho_W/\rho_L$ is lower than $W_{crit}$. The probability to observe such ratio $\rho_W/\rho_L$ which is lower than $W_{crit}$, given that $H_0$ is true, is exactly $\alpha$.
4) The power $\Pi$ of this test is $W_\Pi$ (the power of a test is defined as the probability that $H_0$ is rejected given that it is false). This is the area (in %) under the estimated density curve of $\rho_W/\rho_L$ for lognormal samples for the x-values below $W_{crit}$.

The test $H_0$ : Lognormal versus $H_A$ : Weibull can be performed using a similar technique as presented above. This test defines the values $L_{crit}$ and $L_\Pi$.

As mentioned above, the choice of the significance level should be chosen low. Nevertheless, this level shouldn't be is chosen too low in order to obtain an acceptably high power.

Table 1
Summary of testing $H_0$ : Weibull for the example considered in figure 1.

| | $H_0$ : Weibull | |
|---|---|---|
| $\alpha$ | $W_{crit}$ | $W_\Pi$ |
| 0.05 | 0.993 | 69 |
| 0.1 | 1.000 | 84 |
| 0.2 | 1.009 | 92 |
| 0.4 | 1.022 | 98 |

Table 2
Summary of testing $H_0$ : Lognormal for the example considered in figure 1.

| | $H_0$ : Lognormal | |
|---|---|---|
| $\alpha$ | $L_{crit}$ | $L_\Pi$ |
| 0.05 | 1.014 | 73 |
| 0.1 | 1.007 | 83 |
| 0.2 | 0.999 | 91 |
| 0.4 | 0.989 | 97 |

From the fact that the ratio $\rho_W/\rho_L$ is a pivotal quantity, it follows that the values $W_{crit}$, $W_\Pi$, $L_{crit}$ and $L_\Pi$ only depend on $\alpha$, $n$ and $r$. These critical values and powers were calculated for $n$ = 10,...,128; $r$ = 10,..., $n$ and $\alpha$ = 0.4, 0.2, 0.1 and 0.05. This is done by simulating 80,000 samples for each combination of $n$, $r$ and for the two underlying failure distributions under consideration. The critical values and the powers were obtained using the graphical method presented in figure 2. The accuracy of the calculated critical values is higher than 0.002 where for the powers, an accuracy higher than 2% was attained.

## 6. Illustrative example

In this example, the test proposed in this paper will be applied to the failure times presented in figure 1. For this experiment $n$ = 40 and $r$ = 40. Tables 1 and 2 give the critical values and $\Pi$'s of tests $H_0$ : Weibull and $H_0$ : Lognormal for $n$ = 40 and $r$ = 40 as a function of $\alpha$.

If one wants to find out whether to choose the lognormal or the Weibull distribution, but one has

the strong belief that it is lognormal, testing $H_0$ : Weibull at a rather high $\alpha$-level is advised. Let us for example test $H_0$ : Weibull at an $\alpha$-level of 0.2. This means that 1) if $H_0$ is rejected, the probability that this rejection is wrong, equals 20% (the $\alpha$-level) and 2) if $H_0$ is not rejected, the probability of the wrong decision is 8% (100%-Π). The observed value for $\rho_W/\rho_L$ is 1.011, which means that $H_0$ is not rejected (>1.009). And indeed, the underlying distribution function was...Weibull. Did you make the right choice?

## 7. Conclusions

In this research, a new method is proposed to distinguish between the lognormal and the Weibull distribution. In order to obtain and accurate estimate of $\rho_W/\rho_L$ , the data under consideration should be continuously monitored. The most important assumption is that the data are either lognormal monomodal or Weibull monomodal distributed.

The method has two main advantages. The first and foremost advantage is that it is an easy-to-use method which is based on the principle of probability plotting. And second, the power of the test can always be calculated, which allows the user to calculate the probability of making the wrong decision

The method proposed here is incorporated in a commercially available software package for reliability analysis[6].

## Acknowledgements

The authors would like to thank H. Maes, G. Groeseneken and P. Roussel, members of the Interuniversity Micro-Electronics Centre (IMEC, Leuven, Belgium) for their useful suggestions.

One of the authors (KC) would like to thank the Flemish Science Foundation IWT for financial support.

## References

[1] Chen, C.Y. Wu, M.K. Lee and C.N. Chen, The dielectric reliability of intrinsic $SiO_2$ films thermally grown on a heavily doped Si substrate - characterisation and modelling, IEEE Transactions on Electron Devices, Vol. 34 (1987), pp. 1540-1552

[2] Bain, Statistical analysis of reliability and life-testing models: theory and methods, Marcel Dekker Inc., New York, 1978

[3] Lloyd, On the lognormal distribution of electromigration lifetimes, J. Appl. Phys., Vol. 50 (1979), pp. 5062-5064

[4] Pinto, The effect of barrier layers on the distribution function of interconnect electromigration failures Qual. Reliab. Engin. Int., Vol. 7 (1991), pp. 287-291

[5] LaCombe. and E.L. Parks, The distribution of electromigration failures, IEEE IRPS Proc. (1986), pp. 1-6

[6] Failure '97, Software package for reliability data analysis distributed by DESTIN N.V., Wetenschapspark 1, B-3590 Diepenbeek, Belgium



PERGAMON

Microelectronics Reliability 38 (1998) 1193–1198

MICROELECTRONICS RELIABILITY

# Extended noise analysis - a novel tool for reliability screening

**Gisela Härtler, Ute Golze, Katrin Paschke**

Ferdinand-Braun-Institut für Höchstfrequenztechnik
Rudower Chaussee 5, D-12489 Berlin, Germany

## Abstract

Noise specifies the fluctuations of device characteristics under operation. Low frequency noise can be used to obtain information on the internal device quality under operating conditions. Commonly, only the noise power spectrum is measured in frequency domain. This is equivalent to the second order moments of the fluctuating quantity. Restriction on reliability prediction via power spectrum analysis means restriction on the second order moments of the fluctuating quantity. Our intention is more general. We are searching for tolerance limits of noise characteristics in time and frequency domain capable of separating reliable devices from non-reliable. Extended noise analysis is a new method of noise characterization in time domain. It considers the 2-dimensional distributions of noise signals at the beginning and the end of a time interval. This yields additional information. Quantities like moments of higher order and the parameters of conditional distributions indicate atypical dynamics. The approach was used to characterize the low frequency current noise of InGaAs/AlGaAs laser diodes at currents well below threshold. Failed and surviving items show different distributions of several noise characteristics.
© 1998 Elsevier Science Ltd. All rights reserved.

## 1 Introduction

Low frequency noise is generally expected to be qualified for indicating the reliability of diverse electron devices and failure mechanisms. However, despite many activities [1] results are disappointing up to now. The arguments in favour of noise measurements are: They give an integral characteristic, they are measurable under operation, and variance of random variables usually increases due to the presence of additional factors like defects or irregularities. Arguments against noise measurements: Published results show poor efficiency discriminating between reliable and non-reliable devices [2]. There are obviously internal effects causing increased noise without affecting reliability.

Common noise analysis is based on the second-order moments (variance and covariances) of two-dimensional distributions of the fluctuating quantity in the frequency domain. This information is only sufficient for a zero-mean Gaussian process. Noise measured on defective devices can be biased due to additional parasitic dynamics and the process may deviate from Gaussian. Therefore, we worked out a more general (extended) noise analysis in time domain. Extended noise analysis [3] relies on estimates of the shape of 2-dimensional distributions for classified noise amplitudes. It is based on a time interval of fixed length and visualizes the flow of densities from initial to final states [4] by con-

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00087-0

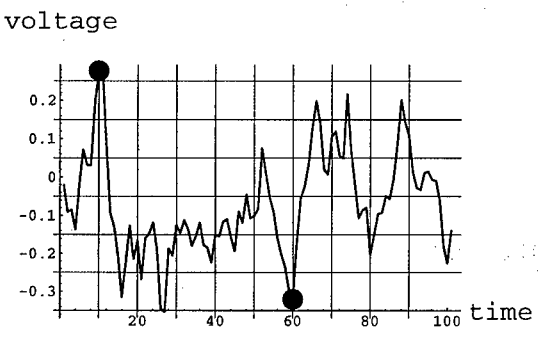



Figure 1: Noise signal transition, 50 time units

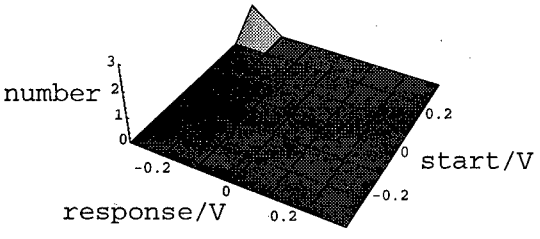



Figure 2: Transition of Fig. 1 plotted into transition count

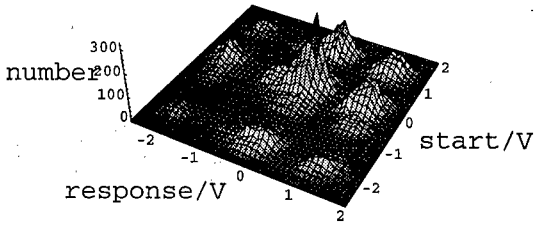



Figure 3: MOSFET-RTS-noise $I_{DS} = 0.4\ \mu A$, $\Delta = 1\ ms$

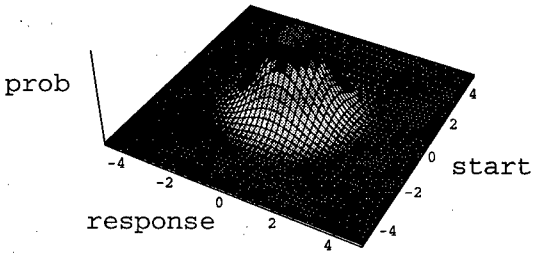



Figure 4: The Gaussian equivalent

ditional probability distributions for the transitions from any initial noise amplitude (start) to all possible final noise amplitudes (response) in the time interval considered. Conditional distributions reveal the stochastical dynamics generating the observed noise process. So it becomes possible to decide whether an observed process is of the expected type or non-homogeneous, mixed, non-linear, etc.

The intention in using noise as reliability indicator is to detect symptoms of failure causing defects. This was done in connection with 1000 h stress tests of InGaAs/AlGaAs laser diodes. The low frequency noise was measured before and after stress tests. Common noise analysis and extended noise analysis were performed. Several noise characteristics calculated from time domain data show differences between groups of surviving and failed laser diodes. We defined tolerance limits for good laser diodes. Although the number of laser diodes is still small a screening with sufficient efficiency seems possible.

This paper gives a short introduction to extended noise analysis. Then the noise characterization of InGaAs/AlGaAs laser diodes will be given. Finally the result of the comparison of failed and surviving laser diodes on the basis of the distributions of some characteristics is shown.

## 2 Extended noise analysis

The noise process $\{X_t\}$ is observed on the discretized scale of noise amplitudes $x$ in discrete time $t$, see Fig. 1. The number of comparable transitions of a signal during a time step of given length is counted and plotted in a 2-dimensional histogram, see Fig. 2.

For a sufficient number of observed transitions the histogram is smooth. It estimates the 2-dimensional distribution behind the second order moments on which the noise power spectrum is based. Following the common assumptions these distributions are Gaussian. Measurements show this is not generally the case. One example of MOSFET noise [5] is given in Fig. 3. Common noise analysis analyzes such a measurement like one single Gaussian process. This distribution is shown in Fig. 4. Defects influencing the operation of electronic devices manifest themselves frequently by non-typical shapes of 2-dimensional distributions.

Conditional distributions are obtained

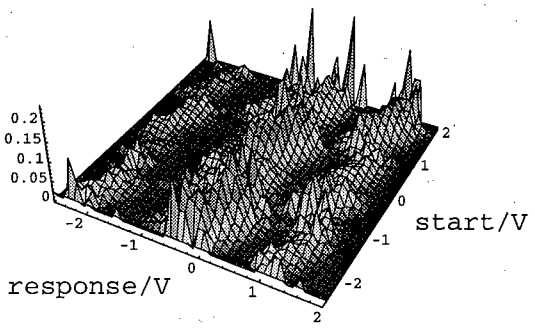



Figure 5: Conditional distributions MOSFET-RTS-noise

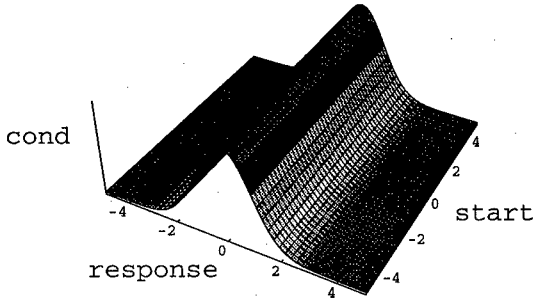



Figure 6: Conditional distributions Gaussian white noise

within each starting-class calculating the relative numbers of transitions into each response-class. One can see whether the conditional averages are on a straight line, conditional variances are equal, or the shapes of distributions are symmetrical and unimodal. Conditional distributions from Figs. 3 and 4 are given in Figs. 5 and 6 respectively.

The set of conditional distributions may be used to characterize the peculiarities of an observed noise process. Without calculation we see the averages roughly as the ridge of modal values (in the case of unimodal and symmetrical distributions). Fig. 7 gives the conditional distributions for a 1 ms time interval of a laser diode failing the test after 41 h. We see the conditional averages depend slightly non-linearly on noise amplitudes at the begin of the time interval. Fig. 8 gives the picture of a surviving laser diode with the same scale. The ridge is rather linear.

From time domain data several noise characteristics were calculated. They may be classified according to the data type as follows:

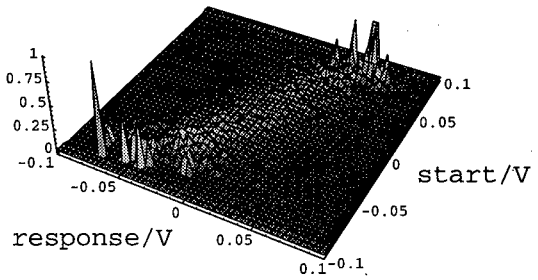



Figure 7: Cond. Distributions, 1 ms laser diode, failing at 41 h

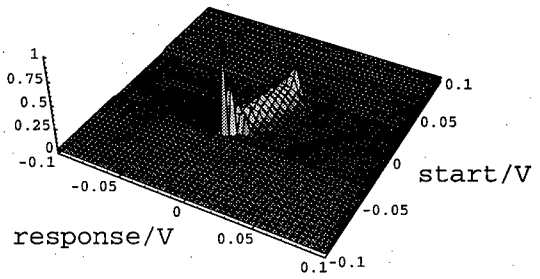



Figure 8: Cond. Distributions, 1 ms laser diode, surviving

1. 1-dimensional distribution of noise amplitudes: mean, variance, skewness and kurtosis.
2. Mixed second order moments: autocorrelation at fixed time interval, time up to a fixed decay of the autocorrelation function.
3. 2-dimensional distributions: mean square departure from linear autoregression, conditional variance compared with the variance of Gaussian model, balance of all variance components including the width of classes, departure from Gaussian distribution, entropy of the 2-dimensional distribution.

Some of the quantities used in extended noise analysis are not common in literature. They are defined as follows:

1. $\tau$ = time until the autocorrelation function (not the autocovariance) decayed below the level $1/e = .3679$.
2. *sqm* = mean square deviation of conditional averages from the straight line. Large values indicate non-linearity.
3. *vnorm* = variance of the mean conditional variance related to the conditional variance

under the hypothesis of a Gaussian process. Large values indicate non-constant conditional variance.

4. *bal* = balance of variance components due to non-linearity and non-constant conditional variance. The ideal value is 0.

# 3 Measurement set-up

This paper is based on the results of 5 different aging experiments executed on InGaAs-QW-GaAs/AlGaAs (experiments 1-4) resp. InGaAs-QW-GaInP/AlGaAs (exp. 5) ridge waveguide (RW) laser diodes ($\lambda$ =1020 nm) made at our institute. For aging conditions and results see Table 1.

Table 1: Conditions of stress tests, total numbers of diodes included and failing

| experiment | T[°C]/P[mW] | n | failure |
|---|---|---|---|
| 1 | 40/200 | 6 | 1 |
| 2 | 40/300 | 9 | 3 |
| 3 | 70/250 | 10 | 2 |
| 4 | 80/250 | 6 | 1 |
| 5 | 40/200 | 10 | 10 |

Noise was measured at currents from 40 nA to 1200 nA well below laser threshold with the aim of detecting surface states considered responsible for facet degradation [6]. Measurements used for the extended noise analysis were made in time domain. For the purpose of comparison additional measurements in frequency domain were carried out. The measurement set-up is given in Fig. 9. To determine the contribution of thermal noise and shot noise to the measurement results a low voltage equivalent circuit for the laser diode was extracted. Fig. 10 shows the noise calculated for the complete measurement set-up in comparison to the measurement. They are in good agreement for high frequencies, up to 300 Hz we found a distinct 1/f-excess noise.

# 4 Reliability prediction for laser diodes

On the basis of the 5 stress tests with the total number of 41 laser diodes included we got

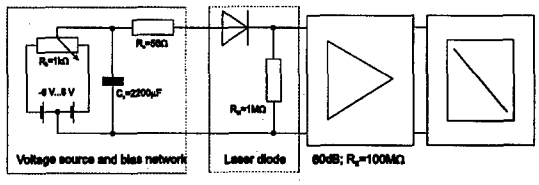



Figure 9: Set-up of noise measurement of laser diodes

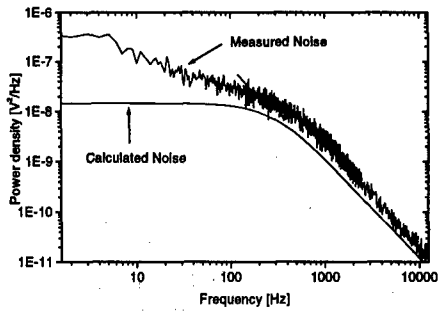



Figure 10: Measured noise compared to calculated noise

24 diodes surviving the tests and 17 failing before 1000 h (comp. Table 1). For all diodes the characteristics given above were calculated. Here 4 selected characteristics calculated from noise measurements at 600 mA will be shown. These are $\tau$, *sqm*, *vnorm*, and *bal*. As reference time step for all calculations the value of $\tau$ was used. This way we ensure to compare two-dimensional distributions of the same correlation.

Figs. 11 to 14 show the cumulative distributions of these characteristics for the samples of failed and surviving laser diodes. The ordinate gives the proportion of characteristics which are smaller than the abszissa value. From Fig. 11 e.g. we see about 80 % of $\tau$'s of unreliable laser diodes don't exceed the maximum value of reliable laser diodes. Screening on the basis merely of $\tau$ could identify about 20 % of unreliable laser diodes since these exceed a certain value. Nonlinearity of conditional averages *sqm*, Fig. 12, allows more efficient screening. On the basis of *sqm* alone we could identify nearly 40 % of unreliable laser diodes. The difference between the distributions of good and unreliable samples indicate the existence of differences in noise characteristics. However, distributions are overlapping and therefore a screening on the basis of only one char-

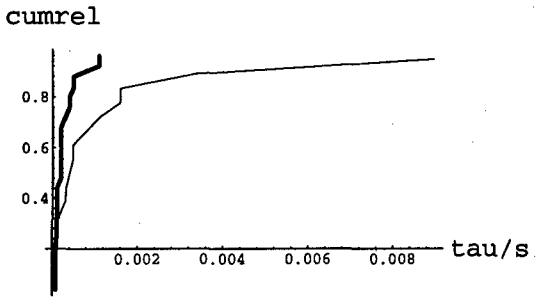



Figure 11: Cumulative distributions of $\tau$, thick line: reliable sample

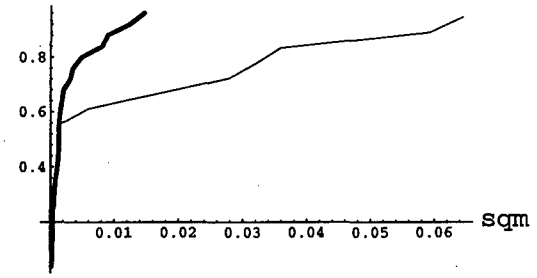



Figure 12: Cumulative distributions of *sqm*, thick line: reliable sample

acteristic is not efficient. An indicator set using 5 characteristics (skewness, kurtosis, minimum and maximum of the standardized variables, balance) and appropriate tolerance limits allowed to recognize all non-reliable laser diodes with 3 misclassifications refusing 3 laser diodes which didn't fail during the 1000 h stress test.

# Conclusions

The intended screening procedure is the result of a learning process. Tests on the basis of 41 laser diodes are certainly not sufficient for optimum selection of screening characteristics and tolerance limits. Taking into mind that the stress tests performed in our institute have the aim of testing new technological variants we are facing a larger variety of defects and failure mechanisms than encountert in routine production with optimized technology. We think that reliability screening on the basis of noise measurements is promising if it is not restricted to

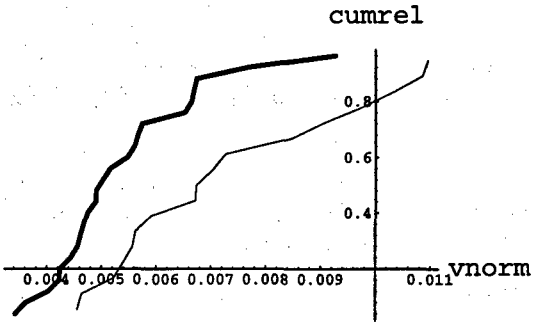



Figure 13: Cumulative distributions of *vnorm*, thick line: reliable sample

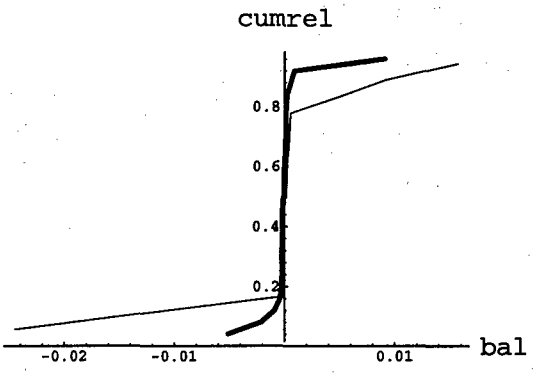



Figure 14: Cumulative distributions of *bal*, thick line: reliable sample

common noise analysis.

# Acknowledgements

The authors thank the Deutsche Forschungsgemeinschaft (DFG) for supporting this work.

# References

[1] Noise as a Diagnostic Tool for the Reliability Improvement of Microelectronic Components for Information Technology and Enviromental Control, NODITO Report, ed. C.Claeys, 1996

[2] Yisong Dai et al., A study of optical noise measurement as a reliability estimation for laser diodes. Microelectron.Reliab. 35, 1995, pp 731-734

[3] Härtler,G., Extended Noise Analysis, in: Noise in Physical Systems and 1/f Fluctuations, eds. C.Claeys, E.Simoen, Proc. of the 14th International Conference, 1997, World Scientific, pp 89-92

[4] Lasota,A., Mackey,M.C., Chaos, Fractals, and Noise, Springer-Verlag, Applied Mathematical Sciences vol. 97, 1994

[5] Härtler,G. et al., Transition Intensities and Noise Spectra in Submicron MOSFETs, in: Noise in Physical Systems and 1/f Fluctuations, eds. C.Claeys, E.Simoen, Proc. of the 14th International Conference, 1997, World Scientific, pp 224-227

[6] Beister,G. et al., Nonradiative current in InGaAs/AlGaAs laser diodes as a measure of facet stability, in press, 1998, Solid-State Electronics

PERGAMON

Microelectronics Reliability 38 (1998) 1199–1210

MICROELECTRONICS RELIABILITY

# Study of degradation mechanisms in compound semiconductor based devices by SEM-cathodoluminescence

G. Salviati*, C. Zanotti-Fregonara*†, M. Borgarino^†, L. Lazzarini*, L. Cattani^†, P. Cova^† and M. Mazzer**

* *MASPEC-CNR Institute, Via Chiavari 18/A, 43100, Parma, Italy*
^ *University of Parma, Department of Engineeringof Information, University Campus, Viale delle Scienze, 43100 Parma, Italy*
** *IME-CNR Institute, University Campus, Via Arnesano 73100, Lecce, Italy*

† *also with Interdepartment Center for Materials and Information Technologies, Viale delle Scienze, 43100, Parma, Italy*

**Abstract**

This papers reports on the microcharacterization of devices for optoelectronic and high-speed electronic applications by spectrally resolved cathodoluminescence. The advantages offered by the high lateral resolution, monochromatic imaging and depth resolved spectral analysis of the technique are presented. In particular, InP based semiconducting optical amplifiers and GaAs based pump lasers for optical fiber communications are characterized from the point of view of compositional inhomogeneities and defect generation in the active regions. GaAs based heterojunction bipolar transistors and high electron mobility transitors are studied to respectively reveal Be outdiffusion from the base and break down walkout after bias aging. GaAs based solar cells are also investigated to show the correlation between dislocations and impurity gettering. Finally the limits of the technique are briefly discussed. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

In recent years there has been a constant trend towards smaller and smaller semiconductor based devices. As a consequence, the need for non destructive techniques with the ability of probing the properties of very small volumes in devices for photonic as well as high speed electronic applications is of great interest. In addition to Photoluminescence (PL), Transmission Electron Microscopy, Chemical Etching etc., also Scanning Electron Microscopy and Scanning Probe Microscopy (SPM), are well established tools for studying both semiconductor materials and devices.

Among all those techniques, Spectrally Resolved Cathodoluminescence (SCL) in the Scanning Electron Microscope (SEM) has reached a certain maturity in the non destructive assessement of optical and electronic properties in both semiconducting heterostructures and devices with high spatial and spectral resolution [1].

As for the near future, semiconductor based devices will use component dimensions of about 100-200 nm and, for instance in the case of quantum devices for photonic and high speed electronic applications, the use of active dimensions from 1 to 10 nm will also be required. It is therefore reasonable to expect that non destructive optical methods for micro-evaluation of degradation

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00128-0

mechanisms in devices will attract more and more interest in the scientific community [2].

This paper will not review the applications of SCL to materials science, due to the number of very good papers and books already published in the field (see for instance [1,2] and references therein enclosed). We will simply report some typical examples of the advantages given by the SCL to the study of degradation mechanisms in semiconductor based lasers and transistors. The results of exploiting the CL high lateral resolution and spectral capability for studying compositional variations and for monochromatic imaging of dark areas (DA) and/or dark line defects (DLD) in active regions of commercial semiconductor optical amplifiers (SOA) and in GaAs based pump lasers for optical fiber communications will be shown.

Further, the advantage of using depth resolved SCL for investigating Be outdiffusion from base to emitter in GaAs based HBTs and for the study of gate-drain breakdown walkout in GaAs based HEMTs will be presented. Finally, the correlation between impurity gettering and internal strain fields in monochromatic imaging of misfit dislocations in GaAs based solar cells will also be discussed.

## 2. Experimental and Results

Cathodoluminescence is the physical process during which a system (in this case a semiconductor) which is in an excited state because of the irradiation from energetic electrons (possibly in an SEM or in a scanning transmission electron microscope), emits photons during the relaxation to a lower energy state. The light emitted by the sample (that can be cooled from room temperature to liquid helium temperature) is collected by a mirror (normally parabolic in shape) and then sent to a monochromator equipped with different gratings and detectors; finally the signal is sent to a computer.

If the temperature is sufficiently low (for instance 20 K for GaAs) excitons (electron-hole-pairs that form a bound state) both free or bound to impurities can be easily studied. In additin to band-to-band transitions, electrons from the conduction band can recombine with neutral acceptors, neutral donors can recombine with holes in the valence band, electrons bound to donors may directly recombine with holes bound to acceptors (at higher impurity concentration) giving rise to donor-acceptor pair (DAP) recombinations [2].

Since the optical emissions occur inside an SEM, there are some advantages with respect to more conventional optical techniques such as photoluminescence and micro-photoluminescence (μPL). In an SEM the electron beam is well defined in energy and can be focused to a very small spot (1-100 nm) with the possibility to scan the sample surface. This results in a map on a submicrometer scale of the intensity variations of the optical transitions on the growth plane. It is worth noticing that, in defect-free or with very low defect density semiconductors, the SCL lateral resolution is mainly limited by the electron-specimen interaction volume (i.e. by the accelerating voltage) [3] and is as low as 100-200 nm, provided the beam diameter is small enough. On the contrary, in case of highly defective samples, the spatial resolution is mainly limited by the minoriry carrier diffusion lenght, due to the efficient non-radiative nature of extended defects like dislocations.

From above, it follows that the first advantage of SCL on conventional PL is the higher lateral resolution. This is interesting for instance in the study of the spatial distribution of optical transitions from quantum emissions from active layers of laser diodes, optical amplifiers etc., (see section 2.1).

The second advantage comes from the possibility of monochromatic imaging of selected emissions with high spatial resolution (see for instance Ref. [4]). This allows, for instance, to distinguish among different mechanisms in the study of the onset and propagation of non-radiative recombination centres like DA and DLD (see sections 2.1 and 2.3).

The third main advantage of SCL is given by the possibility of varying the electron beam penetration depth (i.e. the specimen interaction volume) by changing the accelerating voltage in the SEM. This makes it possible to selectively collect CL spectra presenting optical transitions coming from different zones of the devices both along and perpendicularly to the growth axis [5] (see section 2.2). This is of course true provided that the band gap of the upper layers is larger than the energy of the radiation coming from the layers underneath. An accurate and detailed description of the SCL physical principles, experimental set up and materials applications can be found for instance in Ref. [1,2,6,7]. As for the experimental set up used in this paper, it has already been described elsewhere, for instance in Ref. [8].

### 2.1 InP based SOAs

The integration of semiconducting optical amplifiers with passive waveguides is important for the production of photonic integrated circuits [9]. In these devices, passive waveguides connect the amplifiers, which amplify or switch the optical signals to various optical components. One of the key parameters in the production of those devices is the coupling efficiency between the amplifier and the waveguide. Butt-coupling solid state waveguides is considered a high efficiency matching method that offers flexibility in the amplifier and passive waveguide design compared to other coupling methods like the evanescent [10] and the bundle integrated ones [11]. In addition to the regrowth technique itself, one of the problems that affect the coupling efficiency, is the waveguide layer regrowth layer composition and shape that in turn depend on the growth procedures and on the crystallographic plane of the etched facet of the laser coupled to the waveguide. Here we will show how the CL can detect compositional variations in the regrown InGaAsP passive waveguide and in the InGaAs active layer due to problems occurred during growth and etching procedures.

The base epilayers of the vertical structures of the SOAs studied in this paper have been grown by MOCVD on an InP:S substrate. They consist nominally of 150 nm thick undoped InGaAsP layer emitting at 1260 nm at room temperature, by an undoped multi quantum well active

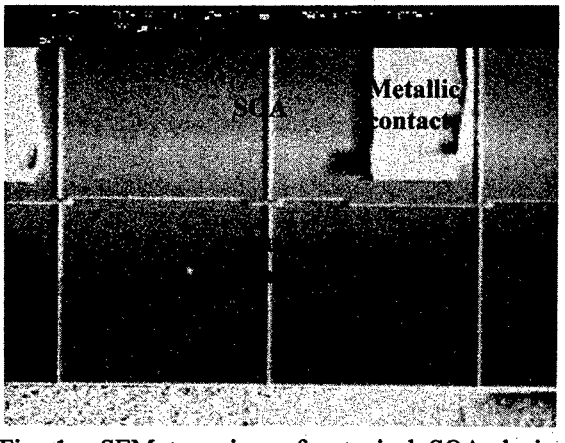



Fig. 1a. SEM top view of a typical SOA device studied.

$In_{53}Ga_{47}As$[well]/$In_{48}Ga_{52}As$(barrier) layer emitting at 1560 nm, by a 150 nm thick InGaAsP layer emitting at 1260 nm, by a $Si_xN_x$ layer followed by deposition of photoresist that is later removed from everywhere except on the ridge. Then the semiconducting structure is attached by reactive ion etching (RIE) and an undoped InGaAsP layer emitting at 1260 nm that is the waveguide, is then regrown by MOCVD. After removing the $Si_xN_x$ layer, a 1300 nm thick InP:Zn clad layer is deposited followed by a final 300 nm thick p+-InGaAs contact layer. Finally, electrical contacts are formed on the top and on the bottom of the wafers. The detailed technological procedures are similar to those reported for instance in Reference [9].

Figures 1a and 1b show two SEM micrographs

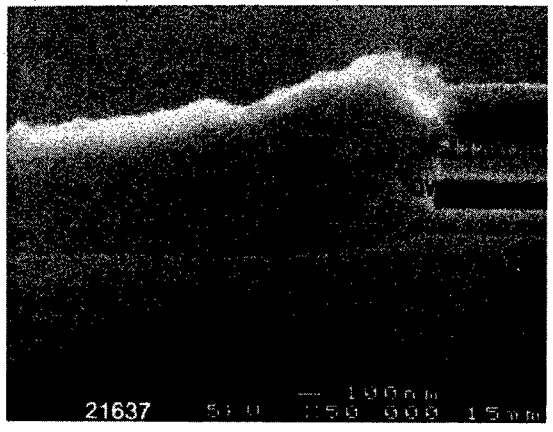



Fig.1b. SEM cross section of the butt coupling vertical geometry.

of the top view and cross section respectively of the SOAs investigated. Figure 2 is an SCL spectrum of the whole area of Figure 1b obtained at 77K. The

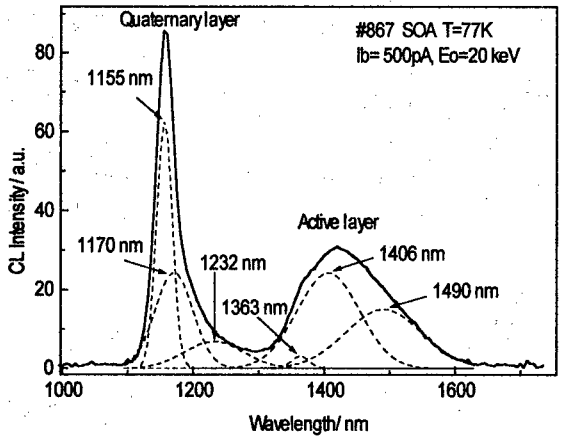



Fig. 2. SCL spectrum of sample in Figure 1.b

main CL peak at about 1155 nm corresponds to the regrown quaternary layer and the broad band centered around 1406 nm corresponds to the main emissions from the MQW active layer. Deconvolution of the band revealed the presence of at least 3 different emissions. Two of them occur at lower energy with respect to the expected nominal one, which should have been at about 1510 nm taking into account the energy shift due to temperature (77 K) and strain.

The MQW has been grown nominally under slightly tensile conditions with an expected emission around 1560 nm at room temperature. The presence of the two additional bands at shorter wavelenght and the large full width at half maximum (FWHM) values of the peaks (116 and 95 nm for the band at 1490 and 1406 nm respectively), suggest that plastic strain relaxation during operation conditions occurred, so increasing the band gap and the FWHM peak values. This has been ascribed to compositional inhomogeneities induced by temperature and/or fluxes instabilities during the growth.

If the peak analysis is correct, extended defects like misfit dislocations should be present at the active layer/confining layer interface. To this purpose, monochromatic CL imaging of the MQW emission was performed. Figure 3 reports a monochromatic CL micrograph of the ridge surface. MDs aligned along the one of the two <110> type directions (vertical black straight lines) are clearly shown. Since the picture has been collected by revealing only the radiation from the main peak of the QW (about 1406 nm), it is clear that the MDs are distributed inside the active layer.

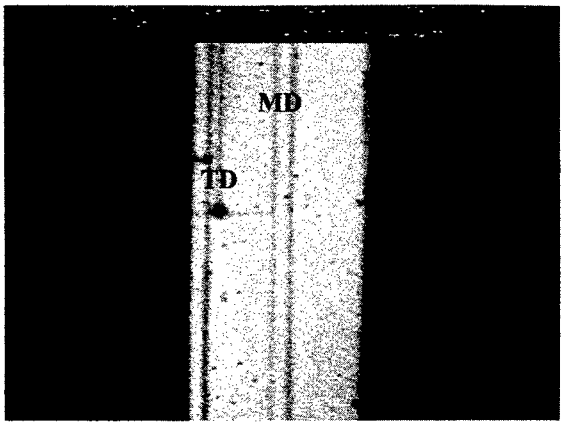



Fig. 3. Monochromatic CL image of MDs inside the QW of a SOA after strain relaxation. 1 cm=15 μm

A remarkably different MD density along the two <110> type directions is also visible. This reveals an asymmetric strain release inside the well that is ascribed to the well known asymmetry of the zincblend lattice [12]. Threading dislocations (TD), most likely at the origin of the MD generation, interacting with MD lines and segments are also shown. Finally, high resolution x-ray diffraction investigations revealed a partial strain relaxation inside the MQW after the regrowth, so fully confirming the CL results.

As reported at the beginning, in addition to compositional variations and plastic deformations in the active region, also the regrowth quality and shape affect the coupling efficiency of SOAs [9]. Figure 4 shows the energy peak position variation in the regrown InGaAsP layer of a different device, moving away from the ridge/regrowth interface. The wavelength emission ranges from 1210 nm at the interface and gets to the expected one (corresponding to the wavelength of the cladding layers, ≈1170 nm) only after 30 μm. On the basis of literature data [13], a possible explanation of this energy variation can be ascribed to the influence of the Si based dielectric mask in perturbing the gas flow dinamics and the surface kinetics of the MOCVD regrowth process. This could result in In and Ga concentration modulation. X-ray microanalysis in the transmission electron microscope (TEM) is necessary to determine which species actually increases or decreases in percentage.

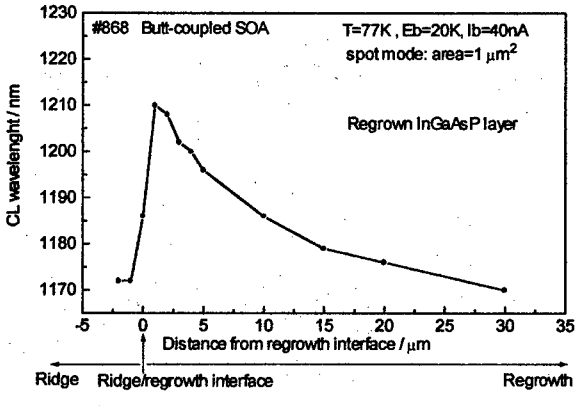



Fig. 4. Compositional variation inside the passive waveguide probably due to the $SiN_x$ mask [13].

The energy shift of about 26 meV roughly corresponds to a compositional variation of about 8-9 %. For the butt coupling SOAs studied in this paper, the wavelength variations are consistent with

those reported in the literature even if the area over which the variation extends is one order of magnitude smaller than those reported [13]. As a secondary effect, an additional strain inside the SOA can be hypothesized as possible additional source of degradation. As a final comment, on the basis of the CL studies, the regrowth conditions have been optimized to give the best optical regrowth quality and coupling efficiency

*2.1.2 GaAs based pump lasers*

High power semiconductor laser emitting at $\lambda \approx 980$ nm wavelength are employed in Er-doped fiber amplifiers (EDFAs) both for telecommunications and single-mode and microcavity lasers [14] to be used in public networks and subsea transmission cables. The reliability issue is a major concern for 980 nm InGaAs/AlGaAs pump lasers. After reducing the frequence of catastrofic optical damage of the laser exit facet by using, for instance, special passivation schemes [15], slow degradation, most likely due to DLDs in AlGaAs, still represents a problem limiting the lifetime which is still much shorter than for InP based lasers [16].

Here we report an example of the advantage offered by CL over the electron beam induced current (EBIC) study of slow degradation mechanisms in a commercial pseudomorphic graded index separate confinement heterostructure (GRIN-SCH) InGaAs/AlGaAs pump laser used in EDFAs and featuring a GaAs/InGaAs/GaAs quantum well (QW) and graded AlGaAs layers. The laser failed after subsea operation for about $10^5$ h in the Mediterranean sea. Figure 5 is an electron beam induced current (EBIC) image of the QW region.

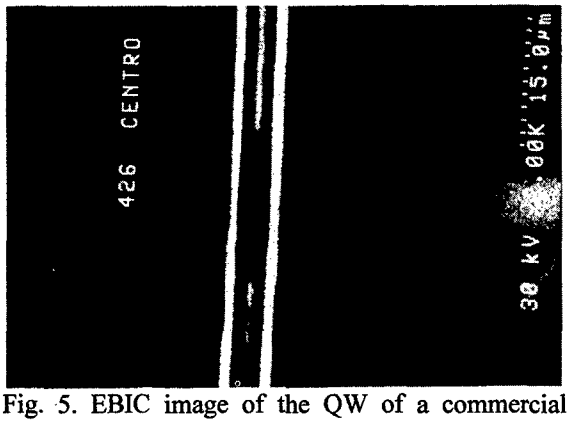



Fig. 5. EBIC image of the QW of a commercial InGaAs/AlGaAs laser employed for under-water transmission. 1 cm = 8 μm

The laser is seen in plan-view. The broken bright line in the centre of the image which is bounded, right and left, by two solid bright lines corresponds to the EBIC signal collected from the current-guiding stripe of the laser. The regions where the line is broken correspond to areas of of reduced EBIC efficiency which occur due to the presence of non-radiative recombination centres like extended defects. In these regions the electron-beam generated electron-hole pairs are not swept towards the contacts by the electric field of the p-n junction laser but are instead captured by the extended defect-states and recombine non-radiatively.

The EBIC technique is extremely sensitive to the presence of non-radiative recombinations [17] in p-n junction devices, however, it lacks wavelength selectivity: this means that the EBIC signal shown in Figure 5 comes from the entire structure and it is not possible, from the EBIC image alone, to determine the precise location of the defects in the vertical structure (i.e. at a depth). SCL may instead be used for this purpose, because each layer in the device emits light at a specific wavelength and therefore it is possible to select the emission from a particular layer. Figure 6 is a monochromatic CL image of the laser shown in Figure 5 taken at a wavelength of $\lambda$=980 nm, corresponding to the peak of emission of the InGaAs/GaAs QW. The laser is observed in plan view, as for Figure 5. A dark area about 30 x 80 $\mu m^2$

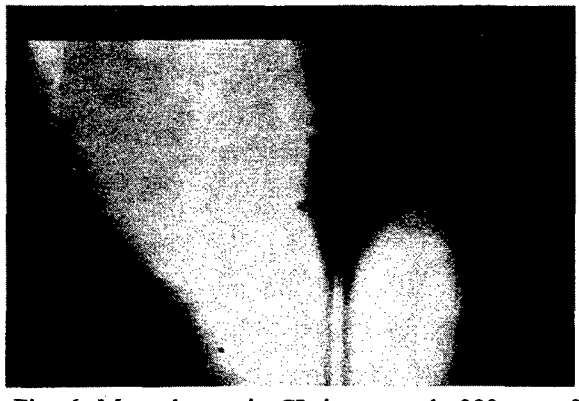


Fig. 6. Monochromatic CL image at $\lambda$=980 nm of the InGaAs/GaAs QW laser, showing CL emission from the active layer. 1 cm = 25 μm

in size is seen in the centre of the image, crossing the stripe and occurs at the same location as the dark area which was observed in the EBIC image (Fig. 5). The dark area shown in Figure 6 therefore confirms that there is significant non-radiative recombination

in the laser. In particular, since this CL image was taken at a wavelength corresponding to the bandgap of the active layer of the laser, it shows that the non-radiative centres are located in the QW.

Fig. 7 is a monochromatic CL image of the laser shown in Fig. 5 taken at a wavelength of $\lambda$ = 680 nm, corresponding to the peak of emission of the AlGaAs layers. A dark area about 20 x 30 $\mu m^2$ is

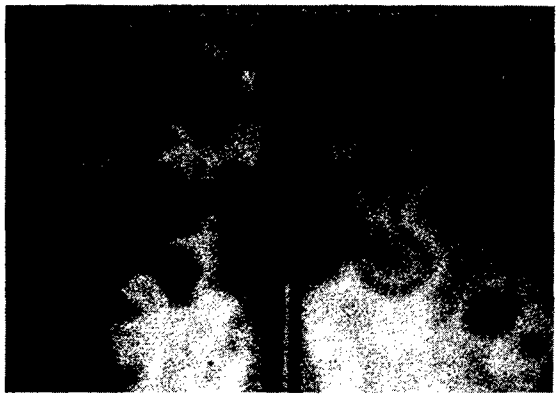


Fig. 7. Monochromatic CL image at $\lambda$=680 nm,of the InGaAs/GaAs SQW laser, showing CL emission from the AlGaAs layers. 1 cm = 10 $\mu$m

present in this image and occurs in the same location as the dark areas shown in Figs. 5,6. The image shows that the non-radiative recombination centres are also present in the AlGaAs layers, although the dimensions of the area over which the defects are present are smaller in the AlGaAs layers, then in the QW. CL imaging cannot distinguish whether the defects may have been generated in the QW and then they may have spread to the surrounding AlGaAs layers, causing device failure, or even simply if the defect density is lower in the AlGaAs layers than in the InGaAs QW.

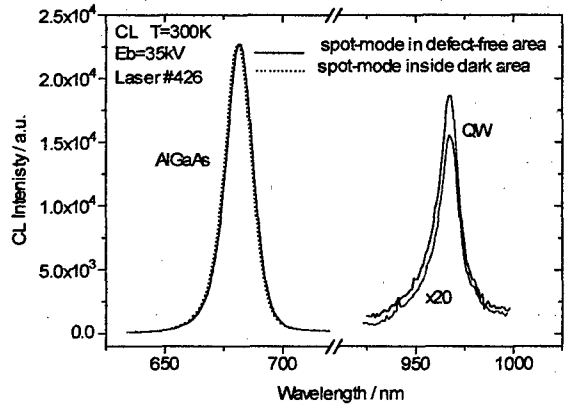



Fig. 8. Plan-view spot-mode CL spectra of the InGaAs/AlGaAs laser.

CL spectral analyses taken in the spot mode inside and outside the dark areas of Figures 6,7 can give some additional information about the origin of the degradation.

Fig. 8 shows plan-view spot-mode CL spectra of the laser shown in Figures 5-7. Two CL emission peaks may be seen at $\lambda$ = 680 nm and $\lambda$ = 980 nm which correspond respectively to the AlGaAs layers and to the InGaAs/GaAs QW. The dotted lines are spectra taken on the current-guiding stripe, inside the dark area shown in Figures 6,7. The solid lines are spectra taken at on the stripe at a distance from the dark area in a region free from extended defects. The spectra show that the defect concentration observed in Figures 6,7 is high enough to cause a significant reduction of the integrated CL emission intensity in the QW material, but not in the AlGaAs layers. The CL results only suggest that a more intense non-radiative recombination occurs in the QW than in the other layers. This can be due to strain relaxation induced by recombination enhanced defect reaction [18] under abnormal operation conditions. Actually, cross sectional TEM studies performed on samples thinned by the focused ion beam techique evidenced <110> oriented MDs inside the QW and threading dislocations, with a much lower density, crossing the AlGaAs confining layers along the {111} planes. This agrees with a recombination enhanced defect glide mechanism suggesting a movement of pre-existing threading dislocations [19, 20].

## 2.2 HBTs and HEMTs

### *2.2.1 Be autdiffusion in GaAs based HBTs*

The ability of the CL to carry out depth resolved investigations of the vertical structures of AlGaAs/GaAs HBTs [HBT-1] has been recently demonstrated [21]. In particular, the CL spectroscopy was applied to investigate the damage induced by the dry etching of the base layer in a double mesa technology and by ion implantation in an implanted technology. CL spectroscopy was used in order to optimize, from the point of view of the ion-induced defects, the HBT fabrication steps. This is important for the production of reliable HBTs because the damage caused by the dry etching used for the fabrication of AlGaAs/GaAs HBTs featuring a half-micron emitter width, induces a quick performance degradatrion during life tests [22].

Beryllium is, together with Carbon, widely employed to p-type dope the base layer of HBTs. It is known that Beryllium easily outdiffuses from the base into the emitter during device operation, giving rise to a degradation of the device performance. Typically, the Beryllium outdiffusion is investigated by Secondary Ions Mass Spectroscopy (SIMS) which, despite its very high analytical sensitivity is a destructive technique. It is therefore interesting to assess the reliability of a non destructve approach to a current problem like dopant diffusion in HBTs.

Here we show the effect on CL spectra of the stress at low current density of base-emitter junctions in AlGaAs/GaAs[Be] HBTs. The comparison between theoretical computations and experimental data indicates that during the stress Be impurities diffuse from the base towards the emitter. In addition, it will be shown how the CL spectra variations induced by the stress can confirm that a REID mechanism is at the root of the Be diffusion.

The heterostructure, grown on a semi-insulating GaAs substrate, has a 500 nm thick GaAs collector Si-doped to Nd=1.5 $10^{16}$ $cm^{-3}$, a 100nm thick GaAs base Be-doped to Na=5 $10^{18}$ $cm^{-3}$, a 10 nm thick undoped GaAs space layer followed by the emitter made by a 20 nm thick AlGaAs layer graded from x=0.0 to x=0.4 Si-doped to Nd=1.5 $10^{17}$ $cm^{-3}$, a 30 nm thick AlGaAs at x=0.4 Si-doped to Nd=1.5 $10^{17}$ $cm^{-3}$, a 20 nm thick AlGaAs layer graded from x=0.4 to x=0.0 and a 400 nm thick GaAs layer Si-doped to Nd=1.5 $10^{17}$ $cm^{-3}$.

The subcollector and emitter contact layers are 500 nm and 100 nm thick GaAs made respectively, doped to Nd=1.5 $10^{18}$ $cm^{-3}$. The subcollector and collector layers were grown by Molecular Beam Epitaxy (MBE), while the layers following the collector were grown by Atomic Layer MBE, in order to minimize the Be diffusion. A detailed description of the growth procedures and device fabrication are reported in Reference [5] .

The HBTs have been electrically stressed by conventional bias aging by forcing through the base-emitter junction a direct current density of about 1000 A $cm^{-2}$ up to four hours at room temperature.
All the CL spectra, performed before and after bias aging stress, were obtained under low injection conditions with a power density of about 1W $cm^{-2}$.
Two sets of CL measurements, carried out on unstressed devices at a temperature of 5K and in the energy range between 1.44 eV and 1.55 eV before and after a bias aging stress, will be preasented and discussed. The CL spectra were recorded by using a Germanium detector. Depth resolved CL investigations were performed at various electron beam accelerating voltages in the range between 14 kV and 10 kV.

Typical CL spectra are shown in Figure 9 for Eo =10 keV. The main peak is at 1.504 eV and there is a shoulder at 1.493 eV. The CL spectra for Eo=14 keV exhibit a main peak at 1.493 eV and a shoulder at 1.504eV. The CL measurements after the stress revealed that the main effect of the stress is an increase of the CL integrated intensity for all the spectra, independent of the accelerating voltage

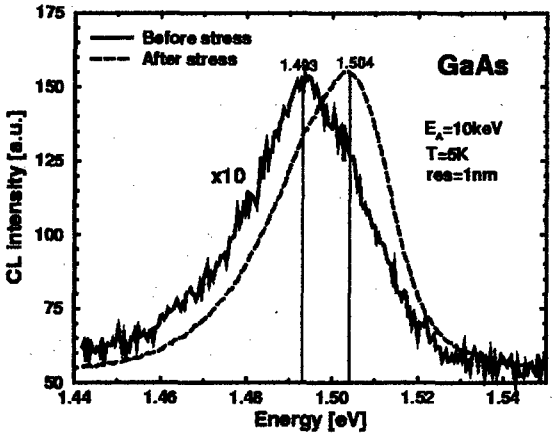



Fig. 9. SCL spectra showing the effect of the Be outdiffusion after bias aging.

range. Note that in Figure 9 the spectrum before stress (solid line) is magnified by a factor 10, in order to be compared with the spectrum after stress (dashed lines). In addition, we can observe that the

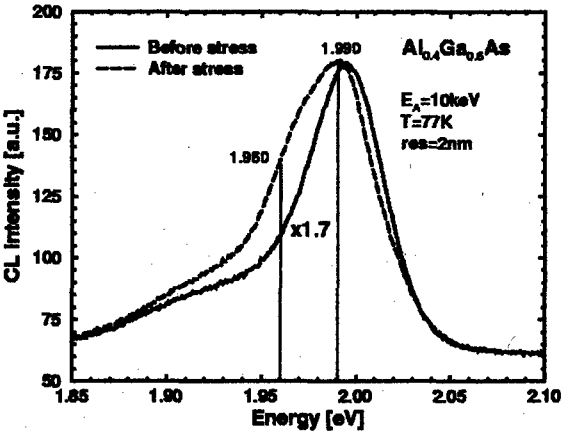



Fig. 10: Typical CL spectra measured at Eo=10keV showing the diffusion of Be inside the emitter.

stress induced a blue shift of the main peak from 1.493 eV to 1.504 eV.

Additional CL measurements on virgin devices at a temperature of 77K in the energy range between 1.85eV and 2.10eV were also carried out. On the basis of the above mentioned results, the CL spectra were collected with a multi-alkaly PM detector. Figure 10 reports the comparison of CL spectra before the stress (solid line) and after the stress (dotted line). After the stress an increase of the CL integrated intensity was observed together with the onset of a shoulder at about 1.960 eV. Note that the spectrum before stress is magnified by a factor 1.7 with respect to that after stress. On the basis of the the values of the optical emissions due to Si and Be in GaAs and AlGaAs reported in the literature [23], we ascribed the emission at about 1.493eV to the Be related transitions in GaAs, the emission at about 1.504 eV to the Si (Nd=1.5 $10^{17}$ $cm^{-3}$) related transitions in GaAs, the emission at about 1.990eV to the Si (Nd=1.5 $10^{17}$ cm-3) related transition in AlGaAs and the emission at about 1.960eV to the Be related transition in AlGaAs. It is worth pointing out that the shoulder at 1.960 eV in Figure 10, is not present in the spectrum before the stress. Due to the stacking sequence of the epitaxial structure of the device, Be impurities should be present only in the GaAs made base layer and the shoulder at 1.960eV demonstrates that during the stress Be atoms diffused towards the emitter, crossing the undoped GaAs spacer.

All the CL measurements revealed an increase of the CL integrated intensity after the stress. This suggests a Be diffusion on the basis of the model reported in Reference [24], that proposed a Recombination Enhanced Impurity Diffusion (REID) mechanism, during which an annhilation of non-radiative recombination centres results in the generation of defects (probably Ga interstitials) boosting the Be diffusion.

The annihilation of non-radiative recombination centers results in an increased CL efficiency after stress as observed experimentally. Single scattering Monte Carlo simulations of electron trajectories confirmed that the CL signal at 10kV comes from the vertical region of the HBT between the base layer and the emitter layer. Since the Be diffusion takes place between the base and the emitter, variations of the spectra can be expected for Eo=10 keV but not for Eo=14 keV as observed experimentally. It must be stressed that the CL results have been confirmed by SIMS analyses [25].

### *2.2.2 Breakdown walkout in GaAs based HEMTs*

Due to the small channel length and very low bandgap, AlGaAs/InGaAs/GaAs pseudomorphic HEMTs [PHEMTs] are very sensitive to hot effects problems [26,27]. In particular, gate-drain breakdown walkout [namely, an increase of the drain-gate breakdown voltage] has been observed in commercial MESFETs, HEMTs and PHEMTs after hot-electron stress [28-30]. This effect was attributed to a build-up of negative charge in the gate-drain region, resulting in a reduction of the peak electric field. We found that cathodoluminescence microscopical images and spectral analyses obtained from unstressed and hot electron-stressed PHEMTs can enable a better insight of the physical mechanisms leading to breakdown walkout in these devices.

The devices under test were delta-doped AlGaAs/InGaAs/GaAs PHEMTs designed for millimeter-wave low-noise applications. The gate length was 0.25 um, the width was 200 um, and the surface is passivated by PECVD-deposited SiN.

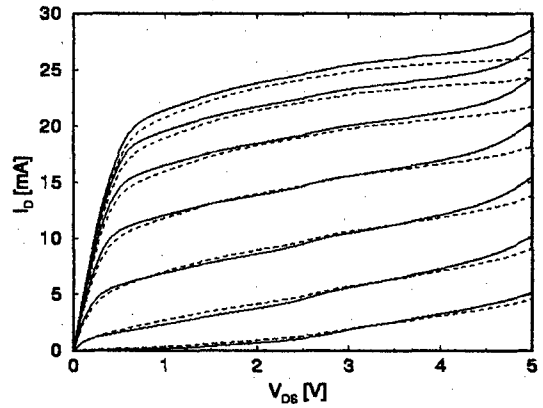



Fig. 11. Output characteristics of one of the PHEMTs before (solid lines) and after (dashed lines) the hot electron stress. $V_{GS}$ ranges from -0.8 V to 0.4 V with 0.2 V steps.

The devices were tested and stressed on wafer by means of a Cascade coplanar probe-station.
The hot electron stress consists of 30 min at $V_{GS}$=0 and $V_{DS}$=4.5 V at RT. Fig. 11 shows the effects of the stress on the DC output characteristics. The most remarkable change is the reduction of the drain current at high $V_{DS}$, where the steeper slope of the pre-stress curves indicates the onset of gate-drain



Fig. 12: 77 K CL top view images, of the gate region of a virgin PHEMT. Top: CL emission; bottom: secondary electron image.

breakdown; after stress this feature nearly disappears, due to breakdown walkout.

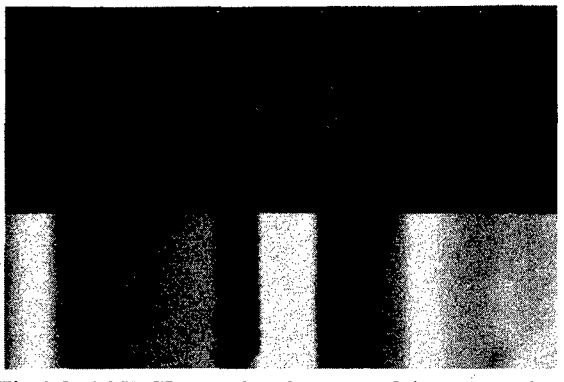


Fig.1 3: 77 K CL top view images, of the gate region of a stressed (b) PHEMT. Top: CL emission; bottom: secondary electron image.

Low temperature SEM CL measurements have been performed before and after the stress with beam energies ranging between 5 and 40 keV. This allowed to perform depth-resolved investigations with high lateral resolution (up to 200 nm). Fig. 12 shows a 77 K CL top view image of the gate region in a virgin device: the signal is brighter at the drain side than at the source for all the devices. The situation is the opposite after the hot electron stress, as illustrated by Fig. 13. 5 K spectral CL analyses confirmed the previous observation.

Figures 14 and. 15 show typical CL spectra on the drain and source of the same sample before and after stress; the reduction of the drain emission after the stress is clearly visible. This observation is in agreement with the mechanism proposed for the breakdown walkout and shown in Figure 11. Hot

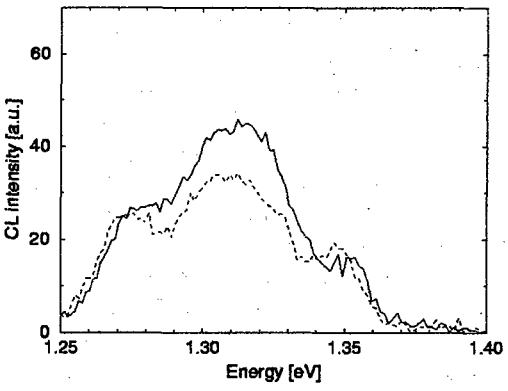



Fig. 14. 5 K CL spectrum from the source (solid line) and drain (dashed line) region in a virgin PHEMT.

electron stress causes the build-up of negative charge between gate and drain, which corresponds to a wider depleted region and then lower peak electric field, both during operation and without biasing (like in CL measurements). Due to the reduction of the electric field the electron-hole pairs generated by the electron beam in CL measurements are less efficiencely separated and than the CL signal is

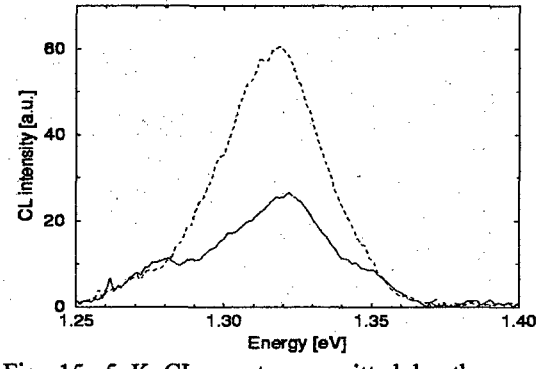



Fig. 15. 5 K CL spectrum emitted by the source (solid line) and drain (dashed line) region in a stressed PHEMT.

higher. In the gate-source region such a signal increase should not be observed, because the charge situation is not altered by hot-electron stress.

### 2.3 GaAs based solar cells

GaAs based solar cells (SCs) currently hold the world record for the efficiency of single junction photo-voltaic cells in unconcentrated (25.1%) and concentrated (27.6%) sunlight and also form the low

band-gap component of the tandem cells with the world's highest efficiency [31]. Further, QW based solar cells have also significant advantages in thermophotovoltaic applications [32]. The use of QW based solar cells improves the efficiency despite the most suitable material for the well, InGaAs, is not lattice matched to the GaAs barrier. This is a major problem since possible plastic lattice deformation induced by strain release during unortodox working conditions results in the generation of MDs that in turn reduce the voltage performance. It is again evident how an imaging echnique with high spatial and spectral resolution like SCL can contribute to the study of the correlation between extended defects and solar cell efficiency. In particular in SCs, in order to understand and control the physical mechanisms at the basis of the generation and propagation of extended defects that act as non radiative recombination centres, is important for improving the light trapping in the QW, the short-circuit current and the device performance. It is well known how impurity gettering at dislocations induces deep levels in the strucutres that can act as traps or recombination sites [32]. This affects the electrical and optical properties of semiconductor based devices, whose mean life time can be strongly




Fig. 16. Monochromatic CL micrograph of a InGaAs/GaAs solar cell at λ=917 nm.

reduced. In this last section we will show how CL not only can evidence impurity gettering at MDs in a MQW based InGaAs/GaAs solar cell [33], but also how it allows also to study the strain field distribution around dislocations. CL micrographs were collected at the main MQW peak and at the satellite peak respectively (not shown here). Typical examples are reported in Figures 16 and 17. In the former (Figure 16), misfit dislocations are visible as dark lines as a consequence of non-radiative recombination dominating the CL contrast. The micrograph of Figure 17 presents quite a different scenario. The luminescence is emitted only by a narrow region around one of the dislocations, the effect of non-radiative recombination being a weak dark contrast in the centre of the bright region. This localised emission is due to an atmosphere of impurities surrounding the dislocation core.

Very important information can be obtained from a monochromatic CL intensity-profiles along a line crossing the dislocation provided that spurious effects due to the non-uniform light collection efficiency along the scan-line are eliminated [34]. By normalising the monochromatic profiles to a reference profile taken near the main MQW peak, not only this problem is cured but all the contributions due to non-radiative recombination are also cancelled out. Fig.18 reports two scan line profiles across the one of the dislocations running vertically in Figure 16 (dotted line) and the

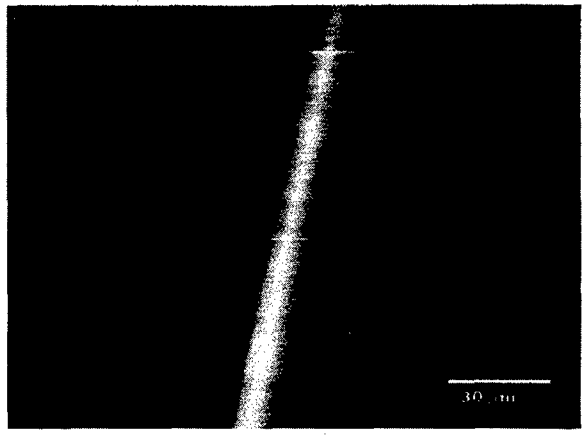


Fig.17. Monochromatic CL imaging at λ=930 nm of impurity gettering around one of the dislocations of Figure 16.

dislocation in Figure 17 (solid line).

Figure 19 shows four normalised profiles collected at four different wavelengths. At 930 nm, that is at the maximum of the satellite peak, the profile is a gaussian centred at the dislocation line, everywhere else the emission is zero. Moving towards the MQW peak, the background signal increases while the satellite peak d ecreases (921.2 nm) until a new maximum appears at about one μm to the left of the original one (918.2 nm). Going beyond the MQW peak, at 916.2 nm, a roughly specular profile is obtained with a maximum at

about 1 μm to the right. This can be explained by considering the strain field generated by the dislocation line. Since the Burgers vector of the misfit dislocation has a component on the growth plane, the associated stress field has an odd

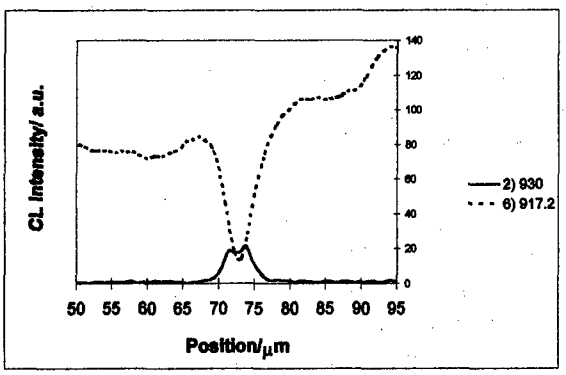



Fig. 18 Scan line profile of CL emission from one of the dislocations in Figure 16 (λ=930 nm) and the dislocation in Figure 17 (λ=917 nm) .

symmetry with respect to the misfit line. This means that at one side of the dislocation the band gap of the emitting material is increased due to the deformation potential while the opposite occurs on the other side.

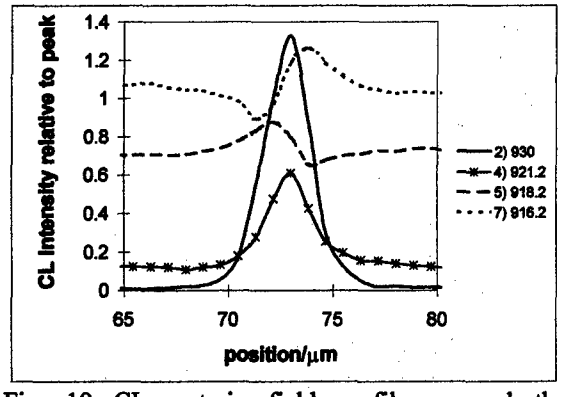



Fig. 19 CL strain field profiles around the dislocation in Figure 17.

Therefore, a detailed map of the stress field around a dislocation can be obtained by CL profiling.

## 3. Conclusions and limits of the SCL technique

Some examples of microcharacterization of semiconducting devices for high speed electronics and optoelectronics by SCL have been presented.

It has been shown how monochromatic imaging and spectral analyses at high lateral resolution can selectively give an insight of the degradation mechanisms in butt-coupled optical amplifiers for photonic integrated circuits, in 980 nm pump lasers for Erbium doped fiber amplifiers and on walkout breakdown in PHEMTs. It has also been demonstrated how the spectral depth resolution of CL can evidence, in a non destructive way, Be outdiffusion in HBTs after bias aging. Finally, the correlation between impurity gettering and strain field distribution around MDs in advanced solar cells has been shown to be possible by SCL profiling.

The examples shown have also evidenced how CL investigations must be complemented by quantitative analytical and structural techniques like SIMS, High Resolution X-ray Diffraction (HRXRD), Rutherford backscattering Spectroscopy (RBS) etc. for the correct comprehension of the origin of failure mechanisms in semiconductor based devices. CL in fact does not often allow to quantify some major parameters as, for instance, the influence of strain release in lattice mismatched heterostructures on SCL peaks energy positions. This cannot be univocally determined unless HRXRD and/or RBS have been carried out in advance. The determination of MDs linear density is possible by CL imaging only at the very beginning of the plastic relaxation when only a few MD lines are present. This is due to the fact that normally for linear MDs densities larger than $10^5$ $cm^{-1}$, at least 5-10 MDs are present inside a sigle black line in CL micrographs; in this case TEM is essential. The quantitative determination of compositional variations with high lateral resolution in QWs and/or confining layers of devices is possible only after HRXRD has given large area average compositional values. Concerning doping and/or impurity density and type determination, again SIMS analyses must be considered as a support of the CL studies etc..

Nevertheless, the non destructivity, the high lateral resolution from the spectral point of view, the monochromatic imaging and the depth resolved spectral resolution are however very interesting for a fast and reliable failure analysis of semiconductor devices.

**Acknowledgements**

Thanks are due to Prof. K.W.J. Barnham from Imperial College of Science Technology and Medicine for providing the solar cells. Thanks are also due to Prof. F. Fantini for very helpful discussions.

## References


[1] B.G.Yacobi and D.B.Holt, J.Appl. Phys., Vol. 59, (1986) R1-24

[2] A.Gustafsson, M.Pistol, L.Montelius, L.Samuelson, J.Appl:Phys, Vol. 84 (2), (1998) R661-765

[3] C.Donolato, Phys. Status Solidi (a) Vol. 141, (1994) k131

[4] D.Bimberg and J.Christen, Inst. Phys. Conf. Ser. Vol. 134, (1993) pp 629-637

[5] F.Fantini, G.Salviati, M.Borgarino, L.Cattani, P.Cova, L.Lazzarini, C.Zanotti-Fregonara, Inst. Phis. Conf. Ser. Vol.160, (1997) pp 503-514

[6] B.G.Yacobi and D.B.Holt, "Cathodoluminescence Microscopy of Inorganic Solids", Plenum Press, NY, (1990)

[7] V.I.Petrov, Phys. Status Solidi (a), Vol. 133, (1992) pp 189-230

[8] C.Zanotti-Fregonara, C.Rigo, A.Stano and G.Salviati, Inst. Phis. Conf. Ser. Vol. 157, (1997) pp 681-684

[9] J.Ahn, K.R.Oh, J.S.Kim, S.W.Lee, H.M.Kim, K.E.Pyun and H.M.Park, IEEE Photon. Technol. Lett., Vol. 8 (2), (1996) pp 1041-1043

[10] D.Wake, S.N.Judge, T.P.Spooner, M.J.Harlow, W.G.Duncan, I.D.Henning, M.J.O'Mahony, Electron. Lett., Vol. 26(15), (1990) pp 1161-1168

[11] M.Renaud, J.A.Cavalles, Ph.Jarry, E.Boucherez, M.Erman, IEEE Photon. Technol. Lett., Vol. 3(1), (1991) pp 47-49

[12] M.S.Abrahams, J:Blanc, C.J.Buiocchi, Appl. Phys. Lett. Vol. 21, (1972) pp 185-186

[13] R.E.Mallard, N.Puets, C.J.Miner, J.Zorzi, D.Adams, M.Cleroux, G.Hillier and R.Moore, Inst. Phys. Conf. Ser. Vol. 146, (1995) pp 591-594

[14] H.Asonen, A.Ovtchinnikov, G.Zhang, J.Näppi, P.Savolainen and M.Pessa, IEEE J. Quant. Electron., Vol.30, N.2, (1994) pp 415-423

[15] A.Moser, A.Oosembrug, E.E.Latta, Th.Forster and M.Gasser, Appl. Phys. Lett. Vol.59, (1991) pp 2642-2644

[16] "New developments in pump lasers", Lasertron, Inc. Commun., Fall (1992), pp 4-5

[17] D.B.Holt and D.C.Joy, "SEM microcharacterization of semiconductors", Academic Press, London (1989)

[18] L.C.Kimerling, Solid Staste Electronics Vo. 21, (1978) pp. 1391-1401

[19] F.Magistrali M.Maini, G.Salmini, A.Bonfiglio, M.B.Casu, M.Vanzi, Proc 23rd Int Symp. For Testing and Failure Analysis ISTFA 97 S Clare CA Oct 1997 pp 189-194

[20] A.Bonfiglio, M.B.Casu, M.Vanzi F.Magistrali, M.Maini, G.Salmini, Proc. Int. Reliability Physics Symposium 1998, Reno, Nevada March 30 - April 2, (1998) pp113-118

[21] A.C.Papadopoulo, C.Dubon-Chevallier, J.F.Brasse, Scanning Microscopy , Vol. 6, no. 1, pp. 97-103 (1992)

[22] K.Mochizuki, S.Isomae, H.Masuda, T.Tanoue, C.Kusano, Jpn. J.Appl. Phys. Vol. 31 (1992) pp.751- 756

[23] L.Pavesi, M.Guzzi, J. Appl. Phys. , Vol. 75, no. 10, (1994), pp. 4779-4842.

[24] M.Uematsu, K.Wada, Appl. Phys. Lett. , Vol. 58, no. 18, (1991), pp 2015-2017.

[25] M.Borgarino, G.Salviati, L.Cattani, L.Lazzarini, C.Zanotti Fregonara, F.Fantini, A.Carnera, in press Journal of Physics D, (1998).

[26] R.Menozzi, M.Borgarino, P.Cova, Y.Baeyens, F.Fantini, Microelectron. Reliab. Vol. 36 (1996) 1899-1902.

[27] C.Canali, P.Cova, E.De Bortoli, F.Fantini, G.Meneghesso, R.Menozzi, E.Zanoni, Proc. IRPS, Las Vegas, USA, (1995) pp. 205-210

[28] P.H.Ladbrooke, S.R. Blight, IEEE Trans. Electron Dev. Vol. 35 (1988) pp. 257-267.

[29] P.C.Chao, M.Shur, M.Y.Kao, B.R.Lee, IEEE Trans. Electron Dev. Vol. 39 (1992) pp. 738-740.

[30] R.Menozzi, P.Cova, F.Fantini, C.Canali, IEEE Trans. Electron Dev. Vol. 43 (1996) pp. 543-546.

[31] M.A.Green et al. Progress in Photovoltaics: Research and Applications, Vol. 6 (1998) pp.35-40

[32] L.Panepinto, U.Zeimer, W.Seifert, M.Seibt, F.Bugge, M.Weyers, W.Schröter, Mat. Sci. Eng. B42, (1996) pp. 77-81

[33] P.Griffin et al. Proc. 14th European Photovoltaic Solar Energy Conf., Barcelona (1997) pp.1732-1740

[34] M.Mazzer, PhD Thesis work, Imperial College of Science Technology and Medicine, University of London, UK (1998)

PERGAMON

Microelectronics Reliability 38 (1998) 1211–1214

MICROELECTRONICS RELIABILITY

# Degradation Behavior in InGaAs/GaAs Strained-Quantum Well Lasers

Tatsuya Takeshita[a], Mitsuru Sugo[a], Teruhiko Nishiya[b], Ryuzou Iga[a], Mitsuo Fukuda[a], Yoshio Itaya[a]

[a] *NTT Opto-electronics Laboratories, 3-1 Morinosato-wakamiya, Atsugi-shi, Kanagawa 243-0198 Japan*

[b] *NTT Access Network Systems Laboratories, 1-6 Nakase Mihama-ku, Chiba-shi, Chiba 261-0023 Japan*

## Abstract

Facet degradation mechanisms of a 980-nm InGaAs/GaAs strained-quantum well laser are analyzed by monitoring the optical-beam induced current. It is clarified that the behavior of defects around the facets govern the long-term stability as well as catastrophic-optical damage generation during operation.

© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

A 980-nm InGaAs/GaAs strained-quantum well laser diode (LD) [1] is a promising pumping source for erbium-doped fiber amplifiers (EDFAs) because of their high gain coefficient and low noise figure [2,3]. Such pumping laser diodes are operated at a high power, which results in catastrophic-optical damage (COD) at the antireflection (AR) coated facet [4,5]. In addition, the severe operating conditions promote long-term LD degradation. These reliability problems usually become a barrier to the system application of such diodes. This paper clarifies the mechanisms of failure and degradation.

## 2. Experiment

Schematic structure of 980-nm InGaAs/GaAs strained-quantum well laser is shown in Fig. 1. The

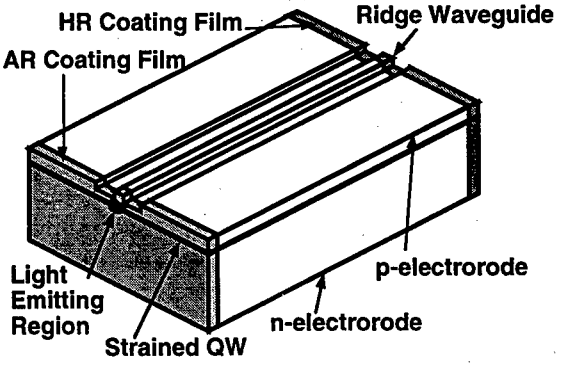



Fig. 1. Schematic structure of 980-nm InGaAs/GaAs strained-quantum well laser with a ridge waveguide.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00144-9

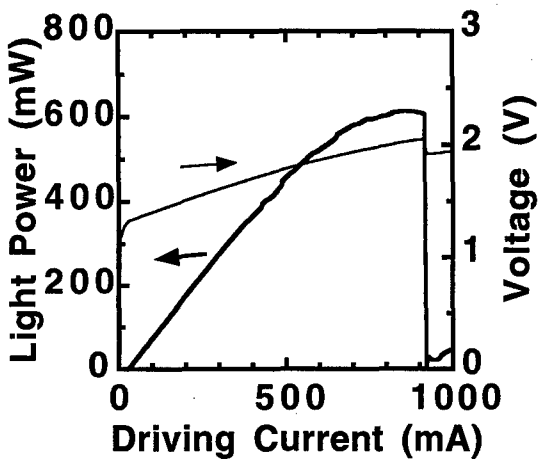



Fig. 2. Typical current and light power characteristics.

laser is a ridge waveguide structure with 2.5 μm width and 900 μm length. The reflectivity of the front facet was lowered to 5% and that of rear facet was increased to 85%. Typical light power and current (L-I) characteristics are shown in Fig. 2. The laser diode shows a sudden reduction in light power caused by COD degradation. A digital optical-beam induced current (OBIC) scanner (JEOL) was utilized to monitor the OBIC at the facet. The light source of the scanner was a 790-nm-laser diode. The incident light power through the objection lens was 0.165 mW, and the estimated spot size was about 1 μm. The penetration depth of the incident light is somewhat less than 1 μm (absorption coefficient: $\alpha > 10^4 cm^{-1}$). A photograph of a digital OBIC scan around the ridge region of a laser diode that failed due to COD generation is shown in Fig. 3. The OBIC intensity is indicated by light and shaded regions; the light region is high intensity. The horizontal light region, as shown in Fig. 3, indicates the pn-junction around the active region of the sample laser diode. The dark spot at the center of the light line corresponds to the active region damaged by COD generation. [6]

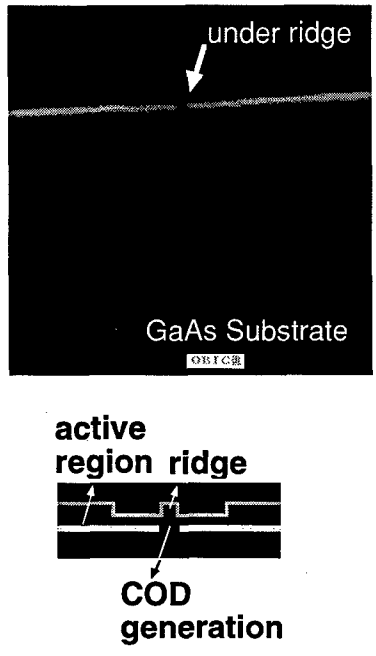



Fig. 3. Photograph of a digital OBIC scan around the ridge region.

## 3. Results and Discussion

The aging test was carried out under a constant output power of 150 mW at 50°C. The current increase ratio as a function of aging time is shown in Fig. 4. The LDs are divided into two groups. The current in-

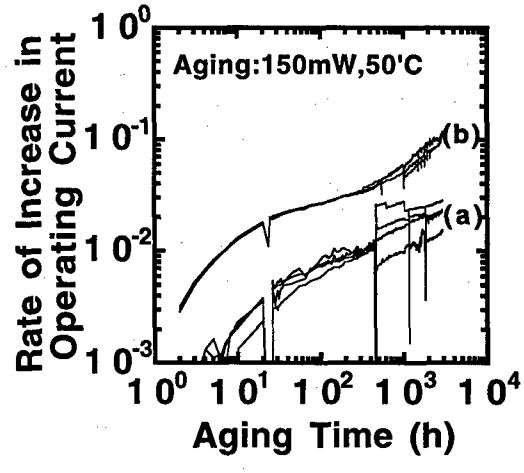



Fig. 4. Current increase ratio as a function of aging time.

crease ratio of group (b) has a high degradation, and is about 10 times as high as that of group (a).

These LDs were periodically taken out of the aging equipment, and the OBIC around the active layer under the ridge as well as the L-I characteristics were non-destructively measured at 25°C. The OBIC of the active region under the ridge was normalized by the OBIC under the other active region. The changes in the normalized OBIC and the maximum power as a function of aging time are shown in Fig. 5. The notations (a) and (b) correspond to groups (a) and (b) in Fig. 4, respectively. The normalized OBIC is also decreased from 0.95 to 0.85 after 3000 hours. The 10% reduction in the normalized OBIC for group (b) is due to the increase in defect density resulting in nonradiative recombination. The maximum power for group (a) is limited by thermal saturation (without COD) and that for group (b) is limited by the COD level. The COD level of group (b) decreased from more than 600 mW to about 170 mW. The OBIC reduction means that the AR facet is degraded and the increased defect density leads to the reduction of the COD level. On the other hand, the normalized OBIC for group (a) does not change for 0 and 3000 hours. The COD level is more than 500 mW for both aging times, which means there is no degradation at the AR facet for group (a).

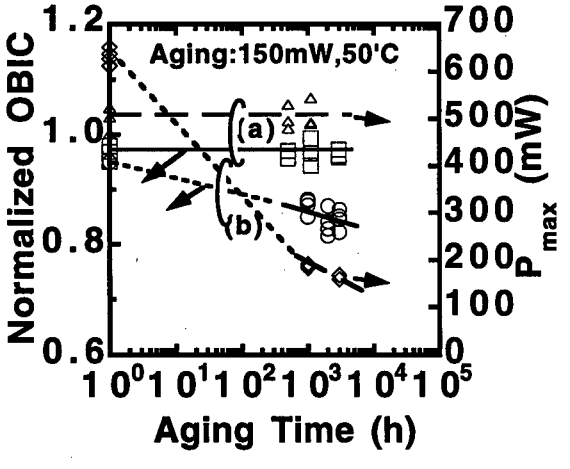



Fig. 5. Changes in normalized OBIC and maximum power as a function of aging time.

The current increase ratio for group (a) was about one order of magnitude lower than that for group (b), as shown in Fig. 4. The normalized OBIC for group (a) scarcely changes during such aging. This means that the large current increase ratio is also determined by the behavior of defects around the facet. The slope of the current increase ratio for group (b) is approximately 0.5 between 1000 and 3000 hours of aging. Based on the reaction model [7], this value means the operating current increase is mainly determined by the diffusion of defects from the facet. Consequently, by reducing the defect density at the facet, long-term stability can be improved along with the suppression of the COD level.

## 4. Conclusion

Facet degradation mechanisms of a 980-nm InGaAs/GaAs strained-quantum well laser have been analyzed by monitoring the optical-beam induced current (OBIC). It has been clarified that the behavior of defects existing around the facet govern the long-term stability as well as COD generation during operation. These degradations are mainly governed by the diffused defects around the facet. The decrease in the density of defects from the facet is important for suppressing catastrophic degradation as well as gradual degradation.

## Acknowledgments

The authors would like to express their appreciation to J. Yoshida, N. Tuzuki, H. Toba, O. Nakajima and M. Naganuma for their helpful advice and encouragement.

## References

[1] Takeshita, T., Okayasu, M., Uehara, S. "High-output power and fundamental transverse mode InGaAs/GaAs strained-layer laser with ridge waveguide structure," Jpn. J. Appl. Phys., vol. 30, pp. 1220-1224, 1991.

[2] Laming, R. L., Farries, M. C., Morkel, P. R., Reekie L., and Payne, D. N. "Efficient pump wavelengths of erbium-doped fibre optical amplifier," Electron. Lett., vol. 25, pp. 12-14, 1989.

[3] Yamada, M., Shimizu, M., Takeshita, T., Okayasu, M., Horiguchi. M., Uehara S., and Sugita, E. "$Er^{3+}$-doped fiber amplifier pumped by 0.98 μm laser diodes," IEEE Photon. Technol. Lett., vol. 1, pp. 422-424, 1989.

[4] Fukuda, M., Okayasu, M., Temmyo, J., and Nakano, J. "Degradation behavior of 0.98-μm strained quantum well InGaAs/AlGaAs lasers under high-power operation," IEEE J. Quantum. Electron., vol. 30, pp. 471-476, 1994.

[5] Moser A., and Latta, E. E. "Arrhenius parameters for the rate process leading to catastrophic damage AlGaAs-GaAs laser facets," J. Appl. Phys., vol. 71, pp. 4848-4853, 1992.

[6] Wallon J., and Devold, P. "Spectral Behavior and Aging Tests of InGaAs/AlGaAs and Al-free 980nm Pump Laser Diodes," ESREF 95, Bordeaux, France, 1995, pp. 415-420.

[7] Fukuda, M., Reliability and Degradation of Semiconductor Lasers and LEDs, Artech House (1991).

PERGAMON

Microelectronics Reliability 38 (1998) 1215–1220

MICROELECTRONICS RELIABILITY

# Early signatures for REDR-based laser degradations

A.Bonfiglio[a], M.B.Casu[a], M.Vanzi[a], R. De Palo[b], F.Magistrali[b], G.Salmini[b]

[a]*University of Cagliari, DIEE-INFM, Italy*
[b]*Pirelli Cavi S.p.A., Milano, Italy*

**Abstract**

The basic theory for Recombination Enhanced Defect Reaction (REDR) as responsible for sudden failures in 980 nm SL SQW InGaAs.pump laser diodes is here tested on experimental constant-current life-test data. A link between the occurrence of long term sudden failures and early detectable signatures has been looked for. For a specific case, where the reacting defects may display saturation effects on their recombination efficiency, a possible screening method may be proposed, able to individuate the great majority of sudden failures at about 1000 hours by inspecting data at 20 hours. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

The occurrence of sudden failures in constant-current life-tested pump laser diodes (980 nm InGaAs SL SQW) has solicited a large effort for finding their origin. A wide set of measurements has depicted a scenario in which native defects evolve from marginal regions, according to the Recombination Enhanced Defect Reaction (REDR) mechanism, and approach the optically active layer. A model for REDR has also been able to describe the time evolution of the degradation, including its "sudden" manifestation after long regular life of the laser diode. EBIC and TEM, then, have exploited the structural characteristics of the ultimate lattice defects, after a catastrophic failure.

The somewhat discouraging result of those studies seems to be, despite of the discovery of the failure mechanism, the absolute impossibility of early screening those devices that will suddenly fail.

Nevertheless, for a set of life-tested specimens, a statistical correlation has been found between some anomalous behaviour after only 20 hours and the final occurrence of sudden catastrophic failures. The paper aims to give account of that case, and to point out how it may be explained within the REDR framework, provided some specific starting situation is assumed, which will be the application domain of a possible early screening.

## 2. The REDR model

A theory has recently been proposed [1] for the growth kinetics of a native point defect located in the neutral region of a laser diode, based on the Recombination Enhanced Defect Reaction mechanism (REDR)[2]. The expected ultimate effect has been shown to be a failure mode with the whole characteristics of an extrinsic mechanism: sudden catastrophic failure, no correlation with any initial measurement, casual-like occurrence.

The theory includes the electrical effects of the growing defect on the current of a laser diode, summarized by the following I(V) formula for the sub-threshold range:

$$I(V) = J_0(V)\left[A + \alpha \exp\left(-\frac{2z_0}{L}\right)\right] \quad (1)$$

Here $J_0(V)$ is the standard Shockley formula for the current density in a pn junction, corrected by the ideality factor $\eta$ for dominating recombination

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00073-0

currents

$$J_0(V) = J_s\left[\exp(\frac{qV}{\eta kT}) - 1\right] \quad (2)$$

A is the active area under current confinement, L is the minority carrier diffusion length in the neutral region that includes the point defect, $z_0$ its distance from the edge of the depletion layer (fig.1) and $\alpha$ is a coefficient, a sort of defect strength, that may be linked to the effective radius of the defect, that is that distance $r_0$ from its center at which the density of excess minority carrier vanishes because of defect-enhanced recombination

$$\alpha = 4\pi r_0 L \exp\left(\frac{r_0}{L}\right) \quad (3)$$

According to the REDR mechanism, the energy supplied at the defect location by carrier recombination allows the defect itself to react. Starting from an ideal isotropic evolution, the more realistic case of dislocation glide along preferential planes leads to a defect evolution [1] well represented by the approaching of the point defect (the head of a defect line) to the depletion layer, according to the following time law:

$$z_d = z_0 + L\ln\left(1 - \frac{v_0 t}{L}\exp\left(-\frac{z_0}{L}\right)\right) \quad (4)$$

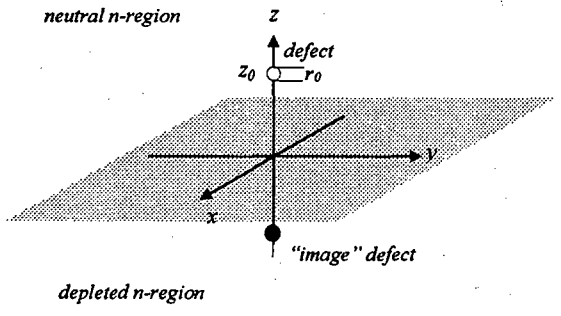



Fig. 1 Localization of the defect in the quasi-neutral region

where the speed $v_0$ is related to material parameters (such as atomic density and energy required to break bonds).

It is straightforward now to collect the indications of that theory into the evolution law for the subthreshold current, that, evaluated for the threshold voltage $V_{th}$ at the junction edges, gives the degradation curve for the most popular laser parameter, the threshold current $I_{th}$:

$$I_{th}(V) = J_{th}(V)\left[A + \alpha\frac{\exp\left(-\frac{2z_0}{L}\right)}{\left(1 - \frac{v_0 t}{L}\exp\left(-\frac{z_0}{L}\right)\right)^2}\right] \quad (5)$$

The effect is then an increasing coefficient for the exponential term of the Shockley law, which could be seen as a *loss of confinement* (increase of A) or, better, an increase of the overall saturation current of the diode [3]. The evolution ends at the time [1]

$$t_0 = \frac{L}{v_0}\left(\exp\left(\frac{z_0}{L}\right) - 1\right) \quad (6)$$

when the defect reaches the boundary of the depletion region, and, at its maximum speed, runs into the active layer.

At that time the fraction multiplied by $\alpha$ attains its maximum value, the unity. Nevertheless, for reasonable deep submicrometric values of the effective defect radius, the actual value for $\alpha$ remains orders of magnitude smaller than the area A, which means the complete undetectability of the phenomenon, by current measurements, at early stages.

As a direct consequence, any current variation detected at early stages (i.e. first steps of a life test) should not be correlated to possible REDR-induced sudden catastrophic failures, that is, no screening should be validated based on early current increases.

## 3. Disagreement with experiments

Actually, early variations of the threshold current $I_{th}$ are often detected in life-tested laser diodes. Fig. 2 reports the experimental data of two 980 nm InGaAs/AlGaAs pump laser diodes under constant current life test and, dashed, the lines of the evolution curves predicted by the above formulas assuming the REDR responsible or of the initial current increase or of the final one of the failed device (A). No match is possible, nor a prediction able to define which of the two fails, based on the initial current change.

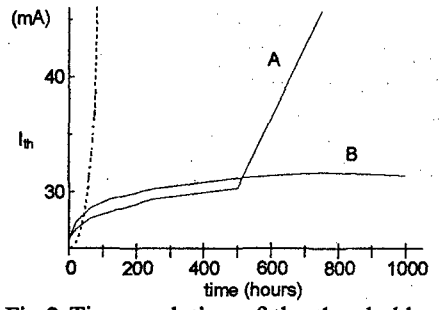



Fig.2 Time evolution of the threshold current $I_{th}$ for two constant-current life-tested laser diodes. No correlation holds between the late strong increase of A and the early variations, similar for A and B (that does not fail). The dashed line draws the expected REDR-driven evolution for the reaction of a single point-defect.

Anyway, EBIC investigations on several specimens (with both A- and B-like evolutions), found statistical correlation between early current increase and extended network of defects (fig.3) under the metal-semiconductor contact [4]. This suggests a possibility for explaining both features: the early increase of the saturation current may be the cumulative effect of many defects (a network, as revealed by EBIC), while the late catastrophic failure is due to the evolution of one, or few, of them. This implies that the great majority of the initial defects will stop their evolution, with very few exceptions that will lead to the failure.

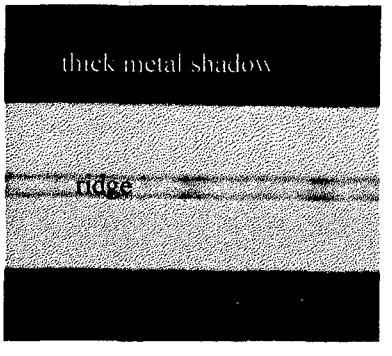



Fig.3 Network of near-surface defects along the laser stripe. Shape, contrast and resolution of the EBIC image allows their classification [4].

This phenomenon is not in contrast with REDR mechanism: many close identical defects, indeed, share a common amount of minority carriers, or, in other words, overlap their carrier destruction ranges (fig.4), which is typically of the order of L. It follows that less carriers, and then less energy, are available for driving the reaction at each defect point. On an experimental ground, TEM confirms (fig.5,[1]) that defects leading to sudden failures are much more localized than the early EBIC-revealed networks and, on the other hand, the above theory easily demonstrates that a planar distribution of identical close defects reduces the reaction energy available for each of them of an amount proportional to the ratio $r_0/L$ between the effective radius of the insulated defect and the minority carrier diffusion length.

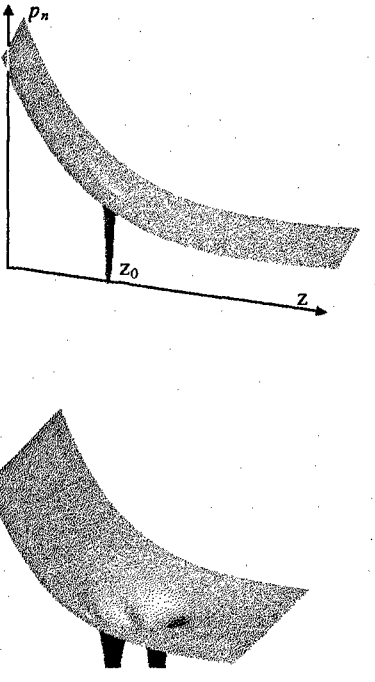



Fig.4 Carrier density near one (a) or two closely spaced (b) point defects. Overlapping of the carrier "sinks" means that less carriers belong to each defect than for the widely-spaced case.

The conclusion, now, is again negative: a lot of initial weak defects will completely hide the presence of more active defects able to continue their REDR-driven motion, up to the device failure.

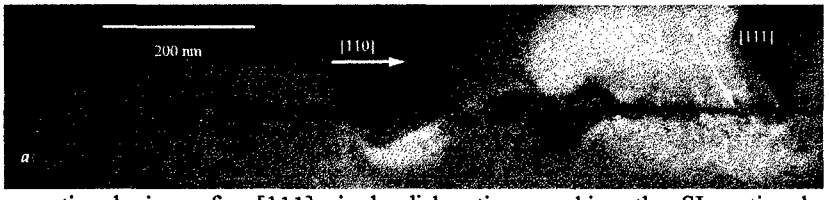



Fig.5. TEM cross-sectional view of a [111] single dislocation reaching the SL active layer, where misfit dislocation patterns originate [1].

## 4. The saturating-defect model

Nevertheless, the situation may change, for the particular initial situation of a limited set of not too closely spaced strong defects. Each of them, indeed, acting at its maximum efficiency, should display saturating effects on the recombination rate, linked to the huge number of excess carriers that should be destroyed inside the small "black hole" for keeping their density null at $r_0$ from the defect point.

Saturation would manifestate as a V-dependent value for $\alpha$. This phenomenon is not observable in the case of many close defects simply because of the reduced amount of available carriers, which ranges well below the saturation limit. In other words, the cumulative effect of many close defects or of some insulated ones should actually differ for the described saturation effect that should take place only for the latter case. If confirmed, a reduction of the defect efficiency should be expected at the higher voltages of the sub-threshold I-V curve, or, in other words, the vertical displacement of the log(I) characteristics should display significant reduction at the V values closest to threshold. Fig. 6 reports the $I_{20}(V)$-$I_0(V)$ differences of the characteristics at 20 hours and 0 hours of two life-tested devices. Dashed straight lines draw the expected slope for pure (non-saturating) Shockley-like excess current. Curve A displays a reduction of the current variation for the higher voltages, following the expectations for saturating defects while curve B completely follows the model of non-saturating defects.

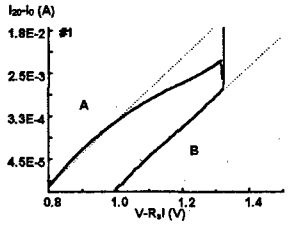
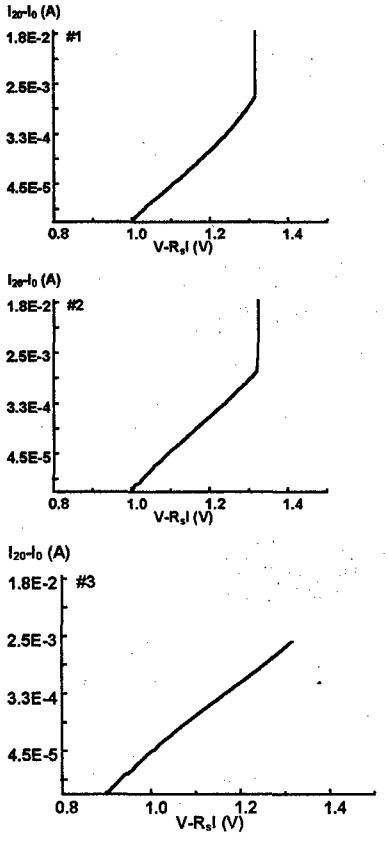



Fig.6 Difference between the 20 hours characteristics and the initial ones for two devices of the life-test.

Fig. 7 summarizes the whole set of early measurements on a group of 14 specimens (#8 accidentally lost). Classification based on the criterion of fig.6 is able to sort the specimens on a saturation basis.

Table 1, on the other hand, reports for each specimen the time-to-failure and their early classification in saturating (S) or non-saturating (NS). Besides, it contains the relative data of the threshold values of current after 20 hrs and at the end of life-tests with respect to the measurement made at 0 hrs. Correlation with the saturation effect of the curves of fig. 7 is evident for the 20 hrs data while it is not for the final ones.



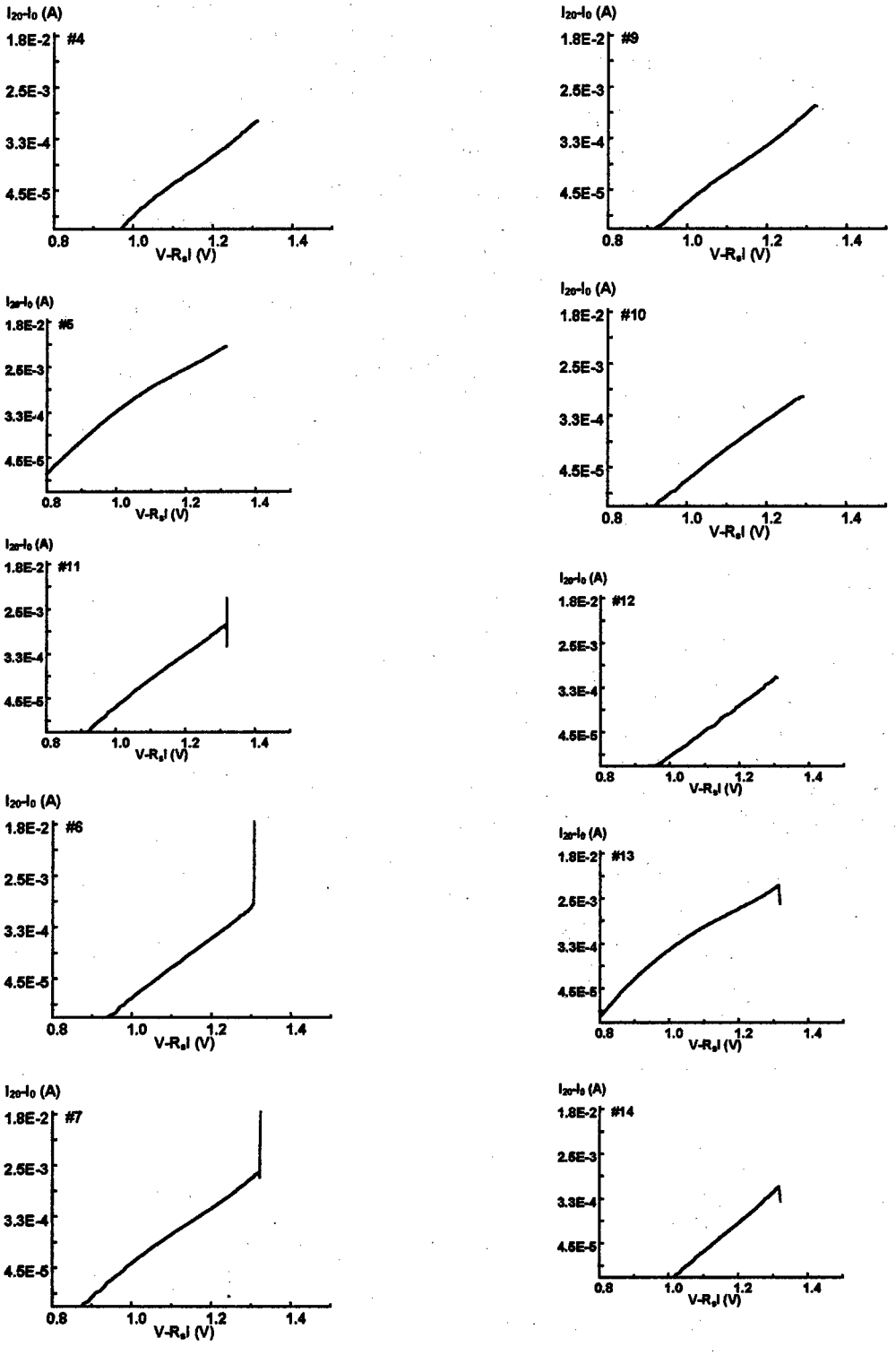



Fig. 7 Early measurements I(20 hrs)-I(0 hrs) vs. reduced voltage of 13 lasers

Table I: Summary of final life-tests data for devices in fig. 7.
S = Saturating NS = Non saturating

| Laser # | Last meas. | Sat. effect | $Ith_0$ (mA) | $\Delta Ith_0$ (mA) ($Ith_{20}$ - $Ith_0$) | $\Delta Ith_f$ (mA) ($Ith_f$-$Ith_0$) |
|---|---|---|---|---|---|
| 5 | 120 | S | 17.3 | 5.5 | 12.3 |
| 13 | 120 | S | 17.4 | 3.3 | 4.2 |
| 7 | 950 | S | 17.8 | 1.4 | 11.0 |
| 1 | final | NS | 17.3 | 0.4 | 1.0 |
| 2 | final | NS | 19.3 | 0.0 | 1.2 |
| 3 | final | NS | 19.1 | 1.6 | 4.4 |
| 4 | final | NS | 16.9 | 0.4 | 2.8 |
| 6 | final | NS | 16.0 | -0.1 | 2.4 |
| 9 | final | NS | 19.2 | 0.5 | 3.3 |
| 10 | final | NS | 13.7 | 3.0 | 6.4 |
| 11 | final | NS | 18.3 | -0.6 | 2.1 |
| 12 | final | NS | 15.6 | 0.0 | 2.6 |
| 14 | final | NS | 16.9 | 0.4 | 0.3 |

## 5. Conclusions

The reported case history, referred to a whole homogeneous set of life-tested devices, has shown statistical correlation between early-stage differential electrical measurements (but not *initial* measurements) and long-term sudden failures. The saturating REDR model may give account of both the observed behavior and the lot-dependent (process fluctuation) possibility of its detectability.
The open question of the original location and structure of the root defects, which could address further theoretical and experimental analysis, could be hopefully investigated by TEM. Localized TEM on root defects, or EBIC mapping of their time evolution would be the most suitable and ultimate methods for the problem of REDR-induced sudden failures in pump laser diodes.

## Acknowledgments

This work has been partially supported by the Italian Istituto Nazionale di Fisica della Materia (INFM).

## References

[1] Bonfiglio A., Casu M.B., Vanzi M., Magistrali F., Maini M., Salmini G. REDR-based kinetics for line defects leading to sudden failures in 980 nm SL SQW InGaAs laser diodes. Proceedings of International Reliability Physics Symposium IRPS 1998, Reno, Nevada, March 30 - April 2, 1998, 113-118.

[2].Kymerling L.C., "Recombination Enhanced Defect Reactions", *Solid State Electronics*, **21**, 1978, 1391-1401.

[3].Bonfiglio A., Casu M.B., Magistrali F., Maini M., Salmini G., Vanzi M., "A different approach to the analysis of data in life-tests of laser diodes", in press on Microelectronics and Reliability, 1998.

[4] Magistrali F., Maini M., Salmini G., Vanzi M., "Analysis of 980 nm SL SQW InGaAs pump laser degradations", Proceedings of International Symposium on Testing and Failure Analysis ISTFA 1995, Santa Clara, California, November 1995, 171-176.

PERGAMON

Microelectronics Reliability 38 (1998) 1221–1226

MICROELECTRONICS RELIABILITY

# Coupling technology impact on low-cost laser modules performances and reliability.

**Marie Morin, Jean-Pierre Defars, Pascal Devoldère**

*France Télécom, CNET Lannion, France*

**Abstract :**

In order to decrease optoelectronic components cost dedicated to the Access Network, new technologies are emerging especially for the optical coupling. Through the investigation of 3 different types of laser diodes modules, a first impact evaluation of coupling technology on both performances and reliability is performed. As far as performances are concerned, especially optical coupled power, the most sensitive devices are receptacle laser diodes. Through temperature cycling, a first evaluation of the long term stability is performed. Again devices based on micro-optics assembly exhibit the weakest behaviour. The most advanced approach combining mounting on Silicon Platform and « detachable pigtail » seems to be the most promising.

© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

To avoid the bottleneck of the introduction of optical fibre into the Access Network, a decrease in optoelectronic components cost is required. To fulfil this requirement, new technologies are emerging, especially for the coupling process which has a strong impact on the overall cost. Meanwhile, these devices will have to face harsh environment, for example wide operating temperature range with the FTTC architecture. Accordingly, the best compromise between low cost, good performances and reliability should be found. In order to evaluate the impact of coupling technology on both performances and reliability, three different types of laser modules have been tested, Receptacle Laser Diodes (RLD's), Laser Diodes in Surface Mountable Package with a detachable pigtail (LDSMP's) and bi-directional modules (BIDI's).

As far as performances are concerned, we have concentrated on Light versus Drive current and on Monitoring photocurrent versus Drive current characteristics (LI curves and Im(I) curves respectively) in the whole operating temperature range (-40 to 85°C) as well as on the related Tracking Error parameter. In addition, the coupled power repeatability has been checked through mating/unmating and full random mating tests for RLD's and LDSMP's. A first evaluation of the reliability, more precisely of the long term coupling stability has been performed through Temperature Cycling.

## 2. Structure of devices tested

A RLD consists in a laser chip in a TOCAN package mounted directly in an optical connector without any pigtail. In these devices, one discrete lens or more focus the light into the patchcord. Those which have been tested are mounted in Fibre Connector/Physical Contact (FC/PC) connectors. A small glass bar, glued onto a self-foc, ensures this Physical Contact [1].

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00099-7

The main feature of LDSMP, as far as optical coupling is concerned is the so called "detachable pigtail". Into the package, a small piece of fibre is soldered on a V grooved Silicon platform, directly in front of the laser chip mounted on the same platform. The other end of this "coupling fibre" is glued in a ferrule and polished. This ferrule forms the optical output of the module which is mated to the detachable pigtail. An important feature of these devices is the package in which there are mounted, an « almost » standard SMT ceramic package as those used in electronic. Associated with the detachable pigtail, this allows to mount these optoelectronic devices as any electronic devices, thus reducing the overall cost.

An other way of saving is to spare one fibre for the network itself by using BIDI. At the present time, most of these devices are based on micro-optics assembly technology with pigtail. A BIDI, consists on an emitter and a receiver in TOCAN packages, a semi-reflective (or dichroïc) beamsplitter, and 2 or more microlens. All elements are dynamically aligned. A BIDI, could be either a duplexer if emitter and receiver work at the same wavelength, or a diplexer if emitting and receiving wavelengths are different.

## 3. Impact on performances

### *3.1 Receptacle laser diodes*

For the tested devices, the most important parameter related to the coupling technology is the coupled optical power and its stability with temperature, especially for RLD and LDSMP devices.

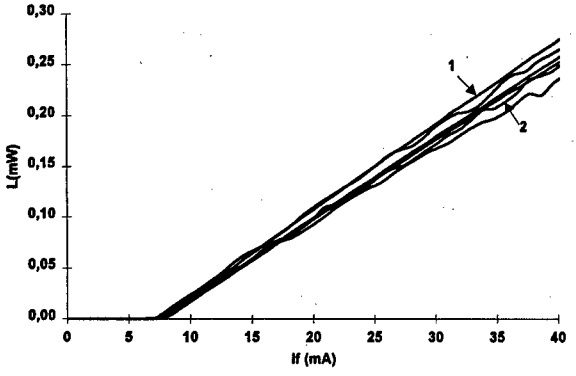



Fig. 1 : L=f($I_{op}$) for different patchcords, FC/PC RLD.

Figure 1 shows typical P(I) curves recorded for a FC/PC RLD connected to 6 different patchcords. Firstly, differential efficiency variation is observed. But, in addition, strong non-linearities are observed all along the P(I) curve for at least 3 patchcords. $I_m$(I) curves which in fact are the rear mirror facet P(I) curves of the laser chip have been simultaneously recorded (figure 2). Again, strong dependence of the differential efficiency with the patchcord connected is observed. For one of them (patchcord 2), the differential efficiency is almost twice the value obtained with the others (figure 2). However, non linearities are not observed along the $I_m$(I) curves, until saturation effect occurred for $I_m$ values close to 600 μA. This saturation effect is clearly intrinsic to the monitoring photodiodes working in photovoltaïc mode and is not related to receptacle behaviour.

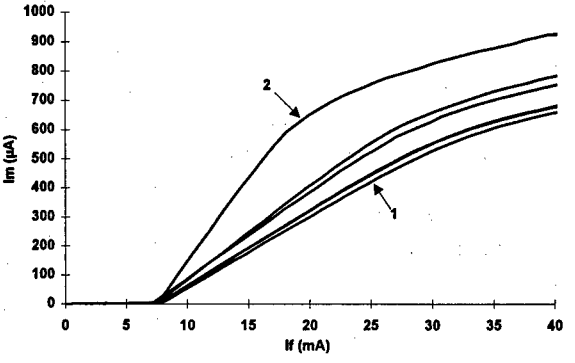



Fig 2 : $I_m$=f($I_{op}$) for different patchcords, FC/PC RLD.

The quite « regular » non linearities observed on the P(I) curves with some patchcords are made more clearly visible by plotting the corresponding dP/dI curve, as shown in Figure 3.

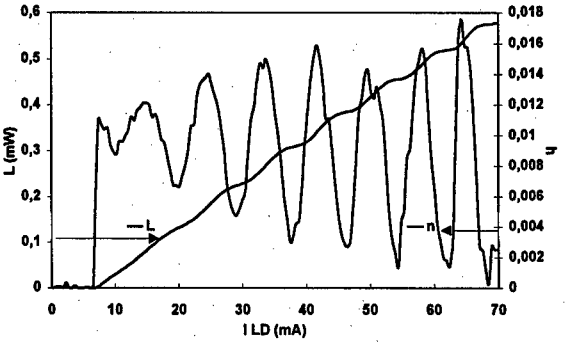



Fig. 3: P and η=dP/dI for RLD N°1 with patchcord 2.

As already underlined, non linearities are not observed on the $I_m$(I) characteristics. Thus, a

preliminary conclusion is that the laser cavity itself is not strongly disturbed by the reflected light. Furthermore, these non linearities disappear when P(I) curves are recorded directly from the FC/PC RLD's with an integrating sphere without any patchcord. This result strongly supports the assumption of an optical cavity formed by refractive index discontinuities within the receptacle but outside the laser chip itself. In addition, one of the « mirrors » of this external cavity should be the optical interface between the ferrule of the patchcord and the small glass bar within the RLD.

Such a behaviour strongly supports the assumption of a rather poor physical contact between the small glass bar of the FC/PC receptacle and the fibre end of the patchcord for all FC/PC RLD's tested.

The LI curves show the impact of the fibre connected to the device on their performances. Usual tests for connectors like repeatability and mating-unmating have been performed to quantify this impact. One example is given with tables 1 and 2 for FC/PC RLD's.

To determine the repeatability parameter, 15 matings and unmatings have been performed consecutively for every device tested, while $I_m$ was maintained at its nominal value. This procedure has been repeated with several patchcords, thus allowing to calculate the random mating. Details of repeatability and random mating calculations have been described elsewhere [1].

Table 1 clearly demonstrates that FC/PC RLD's exhibit scattered values, ranging from 0.07 to 1.34 dB, with strong dependence on the patchcord used for the mating repeatability.

Table 1 : Repeatability values (dB) for 4 FC/PC RLD's.

| RLD n° | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 0.26 | 0.64 | 0.80 | 0.79 |
| 3 | 0.07 | 0.25 | 0.43 | 0.49 |
| 5 | 0.11 | 0.09 | 0.25 | 1.26 |
| 6 | 0.22 | 1.34 | 0.94 | 0.45 |

The random mating test (table 2), leads to even worse results with variation from one patchcord to an other as high as 2.8 dB. More than half of the random mating values are above 1 dB. Of course, the poor behaviour observed with FC/PC RLD's devices at room temperature, is enhanced by changing the operating temperature. Accordingly, it would be difficult to get stable operation in terms of output power with such devices.

Table 2 : full random mating (dB) for 4 FC/PC RLD's.

| RLD no | mating |
|---|---|
| 1 | 0.97 |
| 3 | 2.15 |
| 5 | 2.86 |
| 6 | 2.34 |

### *3.2 Laser Diodes in Surface Mountable Package*

LI characteristics in temperature range have also been performed for LDSMP's as well as mating-unmating test. For these devices, the situation is quite different compared to RLD's. All the LI curves, recorded from the 6 devices tested, exhibit linear characteristics apart from one. As a matter of fact, this only device is showing strange behaviour at some specific temperatures as shown on figure 4. While the LI curves recorded at -40, 25, 45, 65, and 85°C are linear, the -20 and 0°C LI characteristics are strongly sublinear. It should be emphasized that this behaviour has been obtained with this device, whatever the connected detachable pigtail, thus suggesting that such a behaviour is dependent on the device itself. Again, we may attribute the behaviour observed to the physical contact degradation between the « coupling fibre » within the module and the detachable pigtail. As this behaviour is detachable pigtail independent, the most probable assumption is that « coupling fibre » within the LDMSP module has shrunk back thus leading to a microcavity instead of a « true » Physical Contact. This assumption has been confirmed by LDSMP's manufacturer.

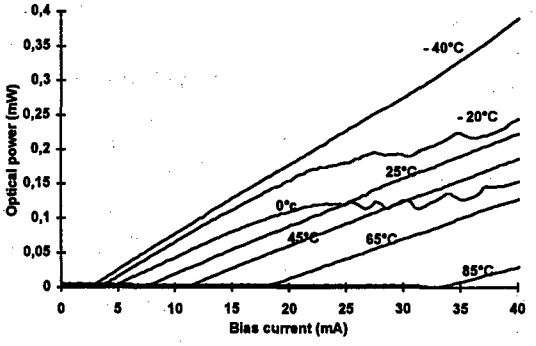



Fig. 4 : L=f($I_{op}$) of the LDSMD with particular behaviour.

As LI characteristics are linear for all devices tested but one and as the « strange behaviour » observed for one device is patchcord independent, the repeatability test has been made for only one LDSMP with 25 mating-unmating. The repeatability figure is +/- 0.05 dB, which is in our measurement precision range. It should be underlined that coupling technologies of RLD's and LDSMP's are quite different. In RLD's, a selfoc and a small glass bar realise the image of the elliptical laser mode at the input of the patchcord fibre in which the optical mode is circular. Due to this modes unsuitability the coupling between the RLD and the patchcord is much more sensitive to small mechanical misalignment, compare to -fibre to fibre- coupling as used in LDSMP's devices.

### 3.3 *Bi-directional modules*

As far as BIDI modules are concerned, all the LI curves recorded do not have any specific feature, whatever the measurement temperature, as it would be expected for devices with « standard » coupling technology (without any connector type interface).

## 4. Coupling stability

To perform a first evaluation of the coupling stability of these devices, and beyond, to check their overall mechanical robustness, temperature cycling has been performed in the -40/+85°C temperature range. The measurement procedure is the following : tested devices are submitted to temperature cycling from $T_{min}$ up to $T_{max}$ and back to $T_{min}$, with a slope of 0.1°C/s. At both extremes $T_{min}$ and $T_{max}$, the temperature is maintained constant for 30 mn. During the whole cycle, devices are operating at their nominal monitoring photocurrent $I_m$, which is maintained constant. All other parameters, P, I, V, and T are continuously recorded. Such a system allows not only to "measure" any drift of optical output power during one cycle, which is the Tracking Error parameter, but also to know the kinetic of any irreversible drift after, for example, 100 cycles.

Coupling stability as a function of temperature range (Tracking Error parameter) is studied for both RLD's and LDSMP's. For the long term stability, BIDI's, based on micro-optics assembly and LDSMP's, which combine Si platform technology and detachable pigtail are compared.

### 4.1 *Coupling stability in the temperature range : Tracking Error*

The precise definition of the Tracking Error parameter Er is as follows [2] :

$$E_{r1} = \max\left|10\log\left[\frac{P_T}{P_{25}}\right]\right| \text{ with } T \in [T_{min}, 25] \quad (1)$$

$$E_{r2} = \max\left|10\log\left[\frac{P_T}{P_{25}}\right]\right| \text{ with } T \in [25, T_{Max}] \quad (2)$$

Where :
$P_T$ is the output power measured at temperature T,
$P_{25}$ is the output power measured at 25°C,
$T_{min}$ is the minimum operating temperature,
$T_{Max}$ is the maximum operating temperature.

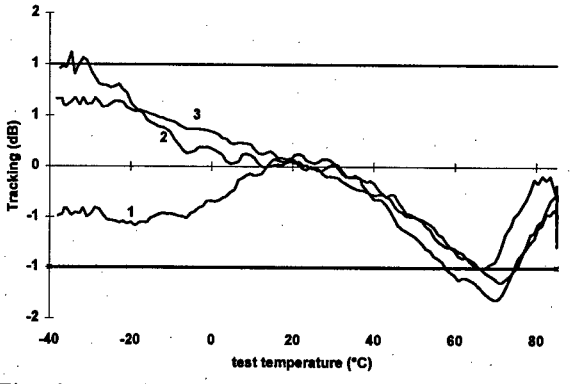



Fig. 5 : Tracking Error for **one** FC/PC RLD with 3 different patchcords.

For FC/PC RLD's, the situation is again very complex. Firstly, the evolution of the Tracking Error with temperature is not at all monotone in the temperature range. Accordingly, to get the maximum Tracking parameter deviation, $E_r$ parameter should be **measured in the whole temperature range** and not only at extreme temperatures as it is commonly done. Secondly, it is difficult to give "absolute" Tracking Error values because they depend on the patchcord mated (as seen in Figure 5 for one RLD with 3 different patchcords). Finally, the values recorded are in many cases above the 1 dB specification.

For LDSMP's, Tracking Error increases almost continuously from $T_{min}$ to $T_{max}$ (Fig. 6) and $E_r$ is in the 1 dB specification apart from device 1. This LDSMP has no linear LI characteristics in the -20/0°C temperature range. As expected, the "abnormal" behaviour also impacts on its Tracking Error parameter in the same temperature range.

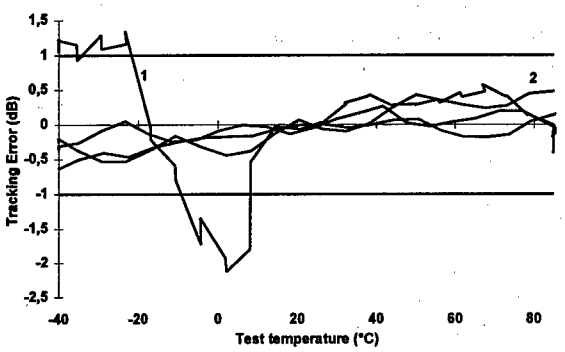



Fig. 6 : Tracking Error for 4 LDSMP's.

Again the two different behaviours observed between RLD's and LDSMP's, as far as Tracking Error parameter is concerned, highlight the impact of coupling technology on performances.

### 4.2 *Long term coupling stability : mechanical robustness*

There are different ways to test the mechanical robustness and the long term stability of a coupling technology. One of them is to perform repeated temperature cycling in the maximum allowed temperature range for the devices concerned. We have performed this test on both BIDI's and LDSMP's components. Temperature cycling conditions were identical to those described in the previous section. (-40/+85°C, slope 0.1°C/s, $I_m$ fixed and continuous recording of P, $I_{op}$, V and T). It should be emphasized that the continuous recording of these different parameters (especially the output power) allows to get the final drift, after 100 cycles, but also the kinetic of the drift, and accordingly allows to know if stabilization is reached or not.

#### 4.2.1 *BIDI's modules*

6 different modules from 3 different manufacturers (2 modules from each), thus 3 different assembly technologies have been submitted to a 100 cycles test. Figure 7 shows the output power drift of those devices at the same temperature of 25°C during the whole test. As can be seen, different behaviours could be observed. For manufacturer M, optical power drift of the 2 devices is rather similar and, lower than -0.5 dB. In addition, the drift occured during the first 10 cycles of the test, and then stabilization is observed. For manufacturers A and S, the behaviour is quite different, with scattered output power drift values ranging from ≅ 0 dB up to -2.5 dB. In addition, for both manufacturers, output power stabilization is not reached at the end of the test.

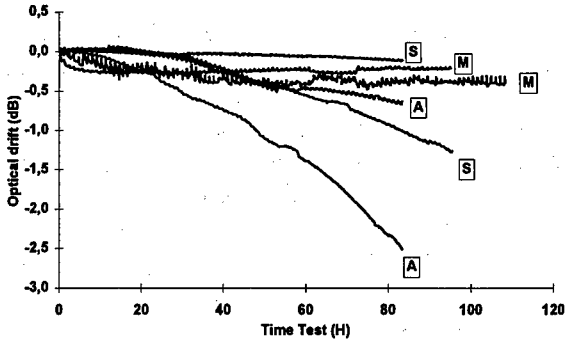



Fig. 7: Optical drift at 25°C for 6 BIDI's during temperature cycling.

These results illustrate that it is quite difficult to ensure long term coupling stability with micro-optic assembly.

#### 4.2.2 *LDSMP's modules*

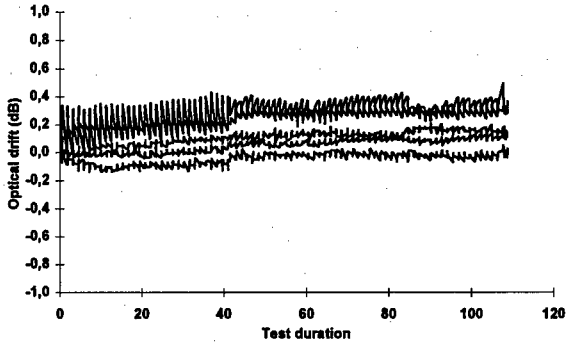



Fig. 8 : Optical stability for 5 LDSMP's during temperature cycling.

On the contrary, the same test performed on LDSMP's devices gives evidence of the good stability of this new coupling technology (Fig. 8). As a matter of fact, none of the modules tested (5 from the same manufacturer) shown an optical power drift higher than 0.4 dB. The smooth oscillations are only due to a small temperature variation in the chamber during the cycling test (maximum 1°C). This demonstrates the stability of the coupling between the laser chip and the "coupling fibre", soldered and glued in a V groove, **but on the same Silicon platform** as the laser chip, thus minimizing differential thermal expansion. This demonstrated as well the physical contact stability between the

"coupling fibre", which is part of the module, and the detachable pigtail.

## 5. Conclusions

In order to decrease the weight of the coupling process on optoelectronic devices cost, several technological approaches are experimented. In return, these coupling technologies may impact both on performances and reliability of the devices concerned. In order to assess such an impact, we have tested emitter modules based on different coupling technologies : laser diodes in receptacle (RLD's), surface mountable modules (LDSMP's), and bidirectional modules (BIDI's). For RLD's devices of the FC/PC type, the optical connection between the receptacle and the patchcord is quite critical, thus leading to poor repeatability of the output power as well as unstable LI curves. In addition, most of the devices based on micro-optic assembly, RLD's and especially BIDI's modules have a rather poor long term output power stability, as demonstrated by the temperature cycling tests. On the other hand, LDSMP's, which combine news technologies such as Si platform and detachable pigtail, exhibit good performances and long term robustness, but again the optical connection between the module itself and the detachable pigtail seems to be the critical point.

## Acknowledgements

Special thanks to M. Gadonna, J. C. Hédé, P. Rochard, J. Bellec for fruitful discussions on optical connectors, and to A. Marie for software assistance.

## References

[1] : Morin M., Térol G., Devoldère P., Some quality issues of low-cost 1300nm Laser Diode packaged-in receptacles, SPIE 13-14 February 1997 San Jose California, volume 3004 pp 71-82.

[2] : International standard, International Electro-technical Commission (IEC), Semiconductor devices : Discrete devices and integrated circuits, part 5 : optoelectronic devices, ref. n0 747-5, 1992-02, pp 110.

PERGAMON

Microelectronics Reliability 38 (1998) 1227–1232

MICROELECTRONICS RELIABILITY

# Failure mechanisms of Schottky gate contact degradation and deep traps creation in AlGaAs/InGaAs PM-HEMTs submitted to accelerated life tests

G. Meneghesso[a], C. Crosato[a], F. Garat[b], G. Martines[c], A. Paccagnella[a], E. Zanoni[a]

a) Universita' di Padova, Dipartimento di Elettronica ed Informatica, via Gradenigo 6/A, 35131 Padova, Italy, IMFM Padova and Consorzio Padova Ricerche.
b) European Space Agency ESTEC, Kaplerlaan 1, P.O.BOX 299, AG Noordwijk ZH, The Netherlands.
c) Universita'di Cagliari, Dip. di Ing. Elettrica ed Elettronica, Piazza d, Armi, 09123 Cagliari, Italy

**Abstract**

In this work we analyze degradation phenomena observed in pseudomorphic AlGaAs/InGaAs HEMTs with Al/Ti gate metallization, which have been submitted to accelerated tests at high drain-source voltage $V_{DS}$ and high power dissipation $P_D$. After these tests, we observe permanent degradation effects, consisting in electron trapping in the gate-drain access region, with consequent decrease in the longitudinal electric field and "breakdown walkout", and in thermally-activated interdiffusion of the Al/Ti gate with decrease in the gate Schottky barrier height and increase in drain saturation current $I_D$. Rather than causing a degradation of the rf characteristics of the device, these phenomena induce an increase in the associated rf gain at 12 GHz, the other rf characteristics being almost unchanged. Overall, the most relevant failure mode observed is an increase of low-frequency transconductance. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

The long term stability in AlGaAs/InGaAs/GaAs pseudomorphic HEMT's (PM-HEMT's) is of crucial importance for their application in the high frequency and low noise application such as satellites and radar. Recoverable degradation has already been observed in commercially available AlGaAs/InGaAs PM-HEMT's [1], [2]. Interaction between impact-ionization phenomena and traps has been pointed out as the responsible for the observed instabilities.

The recoverable drift observed derived from the trapping/de-trapping of hot-electrons, and/or from the recombination of holes generated by impact-ionization with electron trapped in deep levels. Permanent hot-electron induced degradation has been observed in conventional commercially available AlGaAs/GaAs HEMTs [3] and in prototype AlGaAs/InGaAs PM-HEMTs [4]. On the other hand, several failure mechanisms induced by thermally-activated metal-semiconductor interdiffusion have been reported by G. Meneghesso et al. in [5].

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00092-4

In this work we analyze degradation phenomena observed in pseudomorphic AlGaAs/InGaAs HEMTs with Al/Ti gate metallization, which have been submitted to accelerated tests at high drain-source voltage $V_{DS}$ and high power dissipation $P_D$. Main results can be summarized as follows:

(a) Untreated devices are already affected by trapping effects: traps having an activation energy $E_a$=0.4 eV are located in the gate-drain and gate-source access regions; other traps with $E_a$=0.35 eV can be found in the AlGaAs donor layer under the gate and may be identified as DX centers.

(b) The simultaneous presence of these traps and of the impact-ionization phenomena induces a noticeable kink in the output I-V curves [6].

After hot electron and high power dissipation tests, we observe permanent degradation effects, consisting in:

(c) Electron trap generation in the gate-drain access region, with consequent decrease in the longitudinal electric field and "breakdown walkout" [4];

(d) Thermally-activated interdiffusion of the Al/Ti gate with decrease in the gate Schottky barrier height and increase in drain saturation current $I_D$. Rather than causing a degradation of the rf characteristics of the device, these phenomena induce an increase in the associated rf gain at 12 GHz, the other rf characteristics being almost unchanged. Overall, the most relevant failure mode observed is an increase of low-frequency transconductance, especially at low temperature.

## 2. Devices and experimental tests

Tested devices were commercially-available 0.25 μm Al/Ti gate pseudomorphic HEMTs having a gate width W = 200 μm; the AlGaAs donor layer was Si-doped; the gate has a mushroom shape in order to reduce the gate resistance and parasitic capacitance and improve the power handling capability. A complete DC, rf and low frequency characterization has been carried out on untreated devices and at regular times during accelerated tests. Devices have been stressed in different bias points corresponding to different power dissipations and

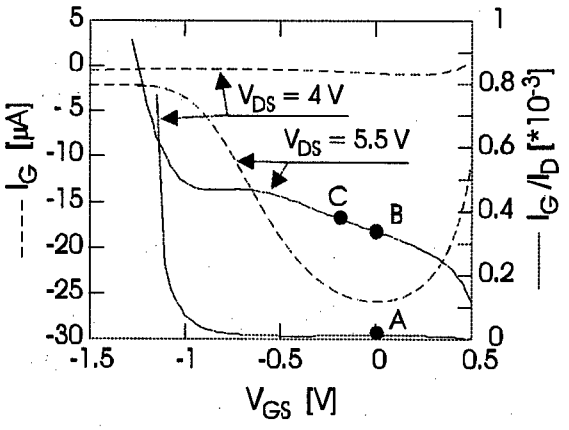



Fig.1. $I_G$ (dashed lines) and $I_G/I_D$ (continuous lines) measured at $V_{DS}$=4 V and 5.5V. The different bias points, adopted for accelerated testing, are indicated.

channel temperatures and to different values of the maximum electric field in the gate-drain access region, the lowest stress level being $V_{DS}$ = 4 V and $V_{GS}$ = 0 V, with minimum power dissipation $P_D$ = 270 mW and channel temperature $T_{CH}$ = 230 °C, hereafter identified as stress "**A**"; other two tests "**B**" and "**C**": "**B**" has $V_{DS}$ = 5.5 V and $V_{GS}$ = 0 V, with $P_D$ = 440 mW and $T_{ch}$ = 355°C. This test has the highest power dissipation and channel temperature. "**C**" was carried out at $V_{DS}$ = 5.5 V and $V_{GS}$ = -0.2 V, with $P_D$ = 360 mW and channel temperature $T_{ch}$ = 300 °C.

Figure 1 shows the gate current, $I_G$, and the gate current-to-drain current ratio, $I_G/I_D$, vs the gate to source voltage, $V_{GS}$, measured in an untreated device at $V_{DS}$ = 4 V and 5.5 V. The typical bell shape behavior of the $I_G$ (due to collection of holes generated by impact-ionization) is almost not visible at low $V_{DS}$ ($V_{DS}$ = 4 V) following the reduction of the electric field. The $I_G/I_D$ ratio, proportional to the multiplication factor in the channel [7], provides an indication of the maximum longitudinal electric field which is the highest for the "**C**" test ($I_G/I_D = 3.8 \cdot 10^{-4}$), lower for the "**B**" test ($I_G/I_D = 3.4 \cdot 10^{-4}$) and the lowest for the "**A**" test ($I_G/I_D = 1.3 \cdot 10^{-5}$).

## 3. Experimental results

Figure 2 shows the output $I_D$ and $I_G$ curves as a function of $V_{DS}$ of a device before (continuous line) and after (dashed lines) the "**B**" stress ($V_{DS}$ =

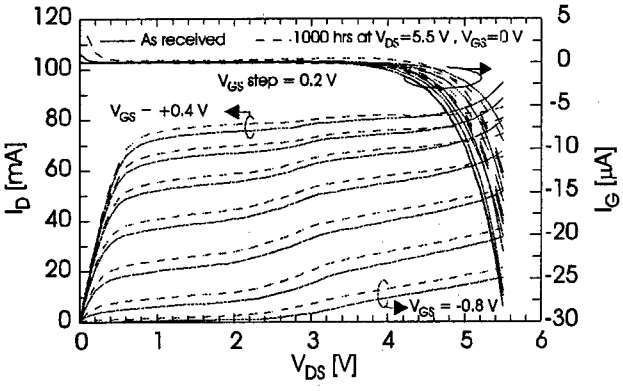



Fig. 2. $I_D$ and $I_G$ vs $V_{DS}$ at different $V_{GS}$ before (continuous lines) and after (dashed lines) accelerated life tests in the "**B**" bias point.

5.5 V and $V_{GS}$= 0 V for 1000 hrs). An increase in $I_D$ after accelerated testing is clearly visible. Moreover, the on-state breakdown voltage increases, as it is testified by the decrease in output conductance which can be noticed at $V_{DS}$ > 4.5 V in open channel conditions.

The gate current $I_G$ due to impact-ionization decreases after accelerated tests, see Fig. 3. This behavior is correlated with the presence of traps in the devices, which can be studied by means of measurements of the frequency dispersion of transconductance, $g_m(f)$ [8].

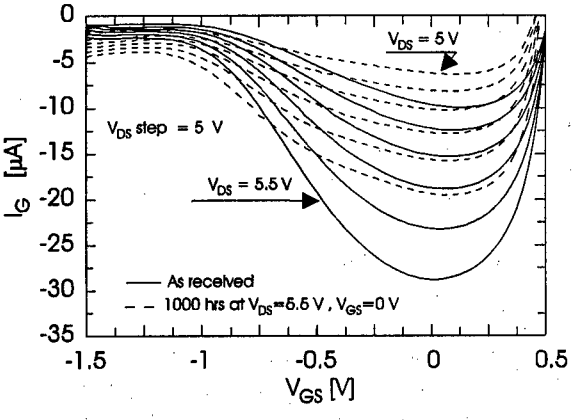



Fig. 3. $I_G$ vs $V_{GS}$ at different $V_{DS}$ before (continuos lines) and after (dashed lines) "**B**" stress test. Decrease of impact ionization is well visible

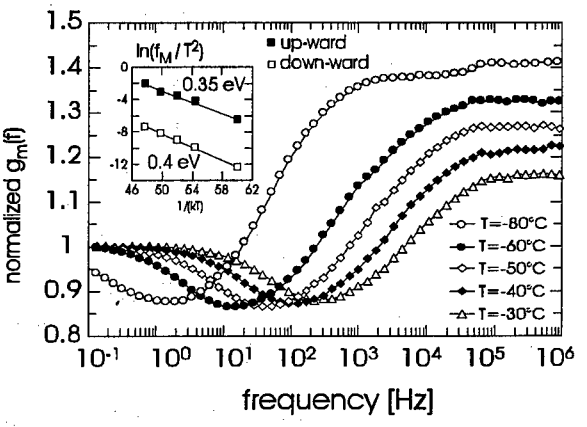



Fig. 4. Normalized $g_m(f)$ measured before stress (bias point: $V_{DS}$=100 mV, $V_{GS}$=+0.4 V). Two deep levels (Ea=0.35 eV and Ea=0.4 eV) have been identified.

The devices present originally both an down-ward and up-ward transconductance frequency dispersion, $g_m(f)$, see Fig. 4. Traps in the access region are responsible for $g_m(f)$ decreasing on increasing frequency [9,10], with Ea=0.4 eV; DX centers in the AlGaAs donor layer under the gate induce $g_m(f)$ increase at increasing f [7]. After accelerated tests, the down-ward peak of $g_m(f)$ increases its amplitude, see Fig. 5, indicating deep level generation in the gate-drain region by hot carriers.

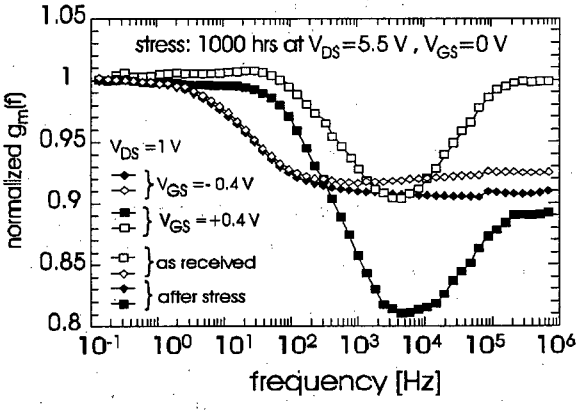



Fig. 5. Normalized $g_m(f)$ before (continuos lines) and after (dashed lines) "**B**" stress test. Increase in the down-ward $g_m(f)$ dispersion is observed in open channel condition ($V_{GS}$=+0.4 V) after stress.

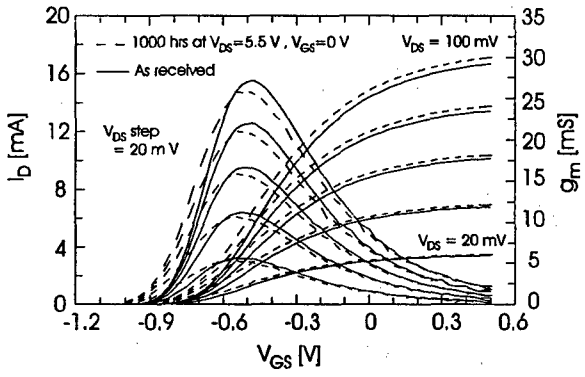



Fig. 6. $I_D$ and $g_m$ vs $V_{GS}$ at different $V_{DS}$ before (continuos lines) and after (dashed lines) "**B**" stress test. Decrease of threshold voltage is observed after stress.

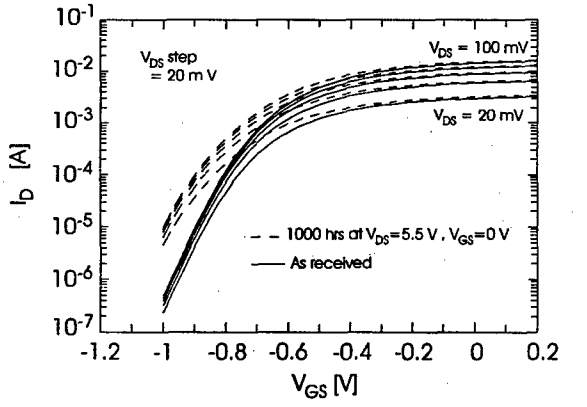



Fig. 7. $I_D$ vs $V_{GS}$ at different $V_{DS}$ in semi-logarithmic scale before and after accelerated tests.

Figure 6 in fact shows the transconductance, $g_m$, measured in linear region, before (continuos lines) and after (dashed lines) "**B**" stress test. A threshold voltage, $V_T$, shift without any significant changing in the $g_m$ amplitude is observed. This suggest that the increase of $I_D$ (see Fig.2) is due to a shift of $V_T$ towards negative values and can not be attributed to any change in the device's parasitic resistances. The shift of $V_T$ is also accompanied by a change in the slope of sub-threshold characteristics, as shown in Fig. 7, which confirm the increase in the deep trap density. The reason for threshold voltage decrease is a corresponding decrease in the Schottky barrier height of the gate contact, $\Delta\Phi_B$, see Fig. 8.

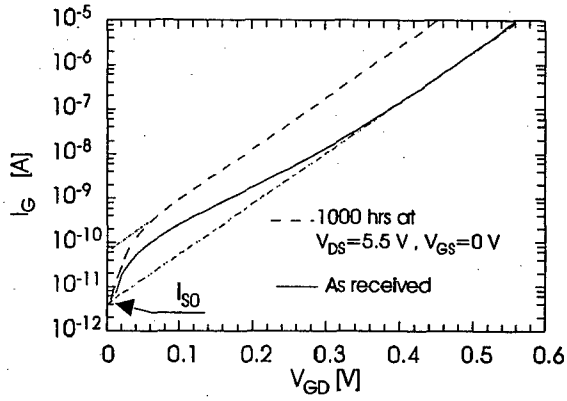



Fig. 8. Gate to drain diode forward current before (continuos lines) and after (dashed lines) "**B**" stress test..

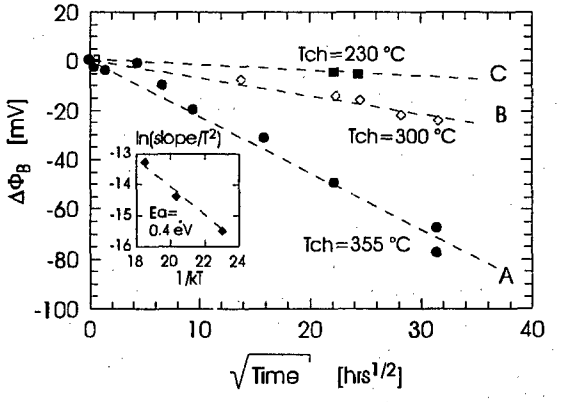



Fig. 9. $\Delta\Phi_B$ vs square root of the stress time and of the channel temperature of the device during stress.

The $\Delta\Phi_B$ is proportional to the square root of the accelerated test time as occurs in a diffusion-limited process, Fig. 9, and is thermally activated with Ea = 0.4 eV. Decrease of Schottky barrier height after thermal treatments has already been observed for Al/Ti Schottky contacts [11].

Figures 10 shows rf associated gain measured at f=12 GHz as a function of $I_D$. A non-negligible increase after tests is observed, indicating a possible decrease of the small signal gate capacitance. We also measured all other relevant rf parameters, without identifying significant changes after the tests. In fact, accelerated testing seems to reduce the dispersion and improve slightly the rf performance of these devices.

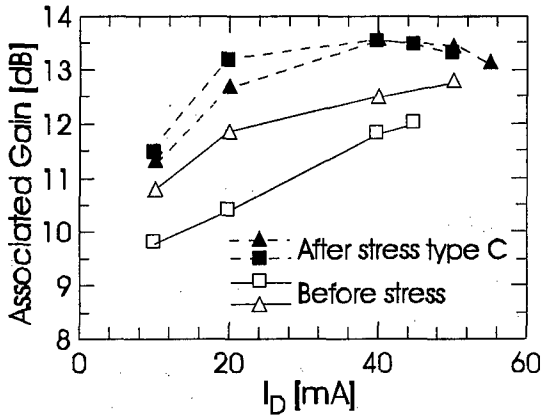



Fig.10. Associated gain (measured at 12 GHz) at different current levels $I_D$, before and after stress.

## 4. Discussion and Conclusions

The degradation phenomena in pseudomorphic AlGaAs/InGaAs HEMTs submitted to hot-electron accelerated tests at high power dissipation $P_D$ have been studied. As received devices present both kink in the output $I_D$ curves and transconductance frequency dispersion. The down-ward dispersion of the $g_m(f)$ has been attributed to traps having an activation energy $E_a$=0.4 eV and located in the gate-drain and gate-source access regions [9,10]. These traps together with pile-up of holes generated by impact ionization (at the source side) are responsible for the kink phenomena [6]. Other traps, with $E_a$=0.35 eV, can be located in the AlGaAs donor layer under the gate, may be identified as DX centers have been identified as responsible for the up-ward $g_m(f)$ dispersion.

After hot electron and high power dissipation tests, we observe permanent degradation effects, consisting in increasing of $I_D$, threshold voltage shift, gate Schottky barrier height decreasing, "breakdown walkout" and increasing in the $g_m(f)$ dispersion.

The thermally-activated interdiffusion of the Al/Ti gate with decrease in the gate Schottky barrier height is responsible for the $V_T$ shift and $I_D$ increase. This mechanism is accelerated by the device self heating.

Traps creation due to hot electrons in the gate-drain access region, with consequent decrease in the longitudinal electric field are responsible for the "breakdown walkout" and for the $g_m(f)$ increase. Trapping of electrons on these levels induces a decrease in the longitudinal electric field, thus explaining the observed reduction in impact-ionization current and breakdown walkout.

Rather than causing a degradation of the rf characteristics of the device, these phenomena induce an increase in the associated rf gain at 12 GHz, the other rf characteristics being almost unchanged.

In conclusions, we have evaluated failure modes and mechanisms of pseudomorphic HEMTs submitted to tests involving the presence of hot carriers and high power dissipation. Drifts in the DC and low-frequency characteristics have been observed, correlated with generation of deep traps by hot electrons and with thermally-activated gate Schottky barrier height decrease.

## Acknowledgement

We want to acknowledge Massimo De Feo (University of Padova) for useful discussion and for setting up the low frequency measurements system. This work has been partially funded by the European Space Agency (ESA/ESTEC) under the contract 162367.

## References

[1] Canali C., Cova p., De Bortoli E., Fantini F., Meneghesso G., Menozzi R. and Zanoni E., Enhancement and degradation of drain current in pseudomorphic AlGaAs/InGaAs HEMT's induced by hot-electrons, Proc. of IRPS 1995, Int. Rel. Phys. Symp. (1995) 205-211.

[2] Meneghesso G., Canali C., Cova P., De Bortoli E. and Zanoni E., Trapped Charge Modulation: A New Cause of Instability in AlGaAs/InGaAs Pseudomorphic HEMT's, IEEE El.Dev. Lett., EDL-17, N.5, (1996) 232--234,

[3] Meneghesso G., Paccagnella A., Haddab Y., Canali C. and Zanoni E., Evidence of interface trap creation by hot electrons in AlGaAs/GaAs HEMT's, Appl. Phys. Lett. Vol. 69, No. 10, (1996) 1411-1413.

[4] Menozzi R., Cova P., Canali C. and , Fantini F., Breakdown walkout in Pseudomorphic HEMT's, IEEE Trans. on El. Dev., ED-43, No. 4, (1996) 543-546.

[5] Meneghesso G., Magistrali F., Sala D., Vanzi M., Canali C. and Zanoni E., Failure Mechanisms due to Metallurgical Interaction in Commercially Available AlGaAs/GaAs and AlGaAs/InGaAs HEMT's, Microel. Rel., Vol. 38, No. 4, (1998), 497-506.

[6] Suemitsum T., Enoki T., Tomizawa M., Shigekawa N. and Ishii Y., Mechanisms and structural dependence of kink phenomena in InAlAs/InGaAs HEMT's, Proc. of IPRM 97, (1997), 365-368.

[7] Hui K., Hu C., George P. and Ko P., "Impact Ionization in GaAs MESFET's", IEEE Electron Device Letters, Vol. 11, No. 3, pp. 113-115, 1990.

[8] Canali C., Magistrali F., Paccagnella A., Sangalli M., Tedesco C. and Zanoni E. Trap-related effects in AlGaAs/GaHs HEMTs IEE Proceedings – Part G, Vol. 138, No.1 (1991) 104-108.

[9] Ladbrooke P. H. and Blight S. R., Low-Field Low Frequency Dispersion of Transconductance in GaAs MESFET's with Implication for Other Rate-Dependent Anomalies, IEEE Trans. on El. Dev. ED-35, No. 3. (1988) 257-267.

[10] Meneghesso G., Paccagnella A., Haddab Y., Canali C. and Zanoni E., Evidence of interface trap creation by hot electrons in AlGaAs/GaAs HEMT's, Appl. Phys. Lett., Vol. 69, No. 10, (1996) 1411-1413.

[11] Wada Y. and Chino K. I., Schottky barrier height variation with metallurgical reaction in Aluminum-Titanium-Galium Arsenide contacts", Solid-State Electr., Vol. 26, (1983) 559-564.

PERGAMON

Microelectronics Reliability 38 (1998) 1233–1237

MICROELECTRONICS RELIABILITY

# Pulsed current stress of Berillium doped AlGaAs/GaAs HBTs

L.Cattani[a], M.Borgarino[a], F.Fantini[a,b]

[a]*DII, University of Parma, Viale delle Scienze, 43100 Parma, Italy*
[b]*DSI, University of Modena, Via Campi 213/B, 41100 Modena, Italy*

**Abstract**

The major concern for reliability of Berillium doped HBTs is the diffusion of the base dopant towards the emitter. This degradation can be enhanced by the device self-heating and/or by REID mechanism. In order to separate the thermal and REID components to the Berillium outdiffusion we performed a pulsed current stress on AlGaAs/GaAs HBTs. In this paper we report on results obtained with different values of the duty cycles for this current. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Heterojunction Bipolar Transistors have received considerable attention for high speed and high frequency applications owing to a number of advantages over the competing devices, in particular due to a better linearity and a higher efficiency at low bias voltage.

The AlGaAs/GaAs HBTs are the devices with the most mature technology, because of the experience accumulated on the AlGaAs/GaAs structure. The Berillium (Be) is the traditional base dopant employed for the fabrication of these devices.

It is well known that, when Berillium is employed as the base dopant, the most serious concern for the device reliability is the outdiffusion of the p-type dopant into the emitter layer during device operation [1, 2]. Nevertheless, Be-doped HBTs have been successfully commercialized [3].

The Be diffusion can be enhanced thermally and/or by the energy released by injected minority carriers recombination (Recombination Enhanced Impurity Diffusion-REID). Since in the power applications the employed currents are high, both the thermal contribution, due to the device self-heating, and the REID contribution [4], due to the large current densities, are important for the Be diffusion. Therefore it's necessary to investigate separately the two contributions. Usually the Be outdiffusion was investigated by means of DC current stress performed at room or high temperature. In order to reduce the thermal component to the Be diffusion sometimes the stresses were carried out at low temperature [5] or on samples back-etched to obtain a lower thermal resistance.

In the present work, for the first time, we have addressed the investigation of the Be diffusion by using a pulsed current stress procedure. In the next sections the device structure and the experimental set-up are described. Then the experimental results are presented in section 5 and discussed in section 6. Finally some conclusion are drawn.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00076-6

## 2. Device structure

The epitaxial structure of the investigated HBTs is reported in Table I. The absence of an undoped spacer and the presence of a graded AlGaAs layer between base and emitter are the most important characteristics for these devices. The wafer was grown by MBE and the devices were fabricated using a standard double-mesa wet-etch process.

All the devices featured multi-finger emitter, a typical layout for power applications.

The device layout is shown in Figure 1: the transistors exhibited four emitters, each one with an area of $10x200\mu^2$.

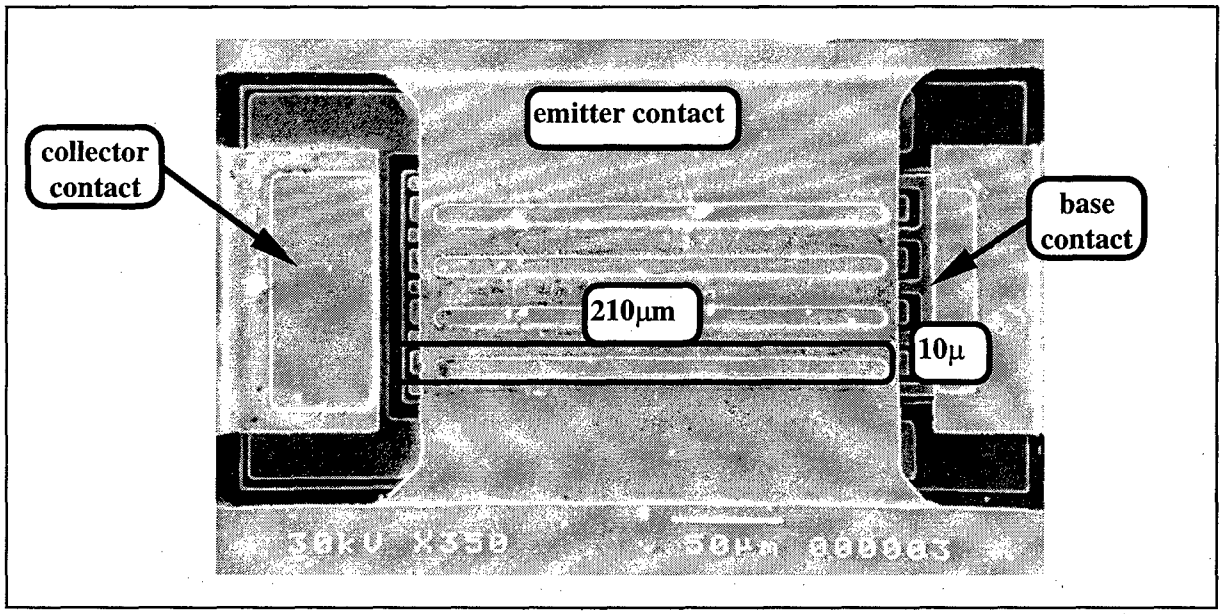



Fig. 1. Layout of the stressed AlGaAs/GaAs HBTs.

Table 1
Epitaxial structure of the stressed HBTs

| | | | |
|---|---|---|---|
| 300 nm | GaAs cap layer | Si | $4.0x10^{18}$ |
| 50 nm | graded (0.28-0.00) | Si | $1.2x10^{18}$ |
| 200 nm | AlGaAs (0.28) emitter | Si | $5.0x10^{17}$ |
| 50 nm | graded (0.00-0.28) | Si | $5.0x10^{17}$ |
| 90 nm | GaAs base | Be | $5.0x10^{19}$ |
| 1000 nm | GaAs collector | Si | $1.7x10^{16}$ |
| 1000 nm | GaAs subcollector | Si | $4.0x10^{18}$ |
| SI GaAs Substrate | | | |

## 3. Experimental details

The devices were DC characterized before the stress procedure, for which a special circuit was designed and fabricated. This stress circuit applyes to the device under test a pulsed current, whose intensity, period and duty-cycle can be regulated.

The four different types of stress are detailed in Table II. The stress 1 was carried out with a duty cycle of 3% and a current of 0.5A, which corresponds to a current density (Jc) of $6.25*10^3$ A/cm$^2$ for each emitter finger. In stress 2 the duty cycle was the same, while the current was increased to 1A (Jc=$13.0*10^3$ A/cm$^2$). For the stresses 3 and 4

the duty cycles were 10% and 20%, respectively. In both cases the stress current was kept at 0.5A.

All the stresses were performed at room temperature and with a period of 100μsec.

The base-emitter junction threshold voltage is known to be the most sensitive parameter to the base dopant outdiffusion [6].

Table 2
Types of stress applyed in this study.

| Stress | Duty Cycle | Jc (A/cm$^2$) |
|---|---|---|
| 1 | 3% | 6.25*10$^3$ |
| 2 | 3% | 13.0*10$^3$ |
| 3 | 10% | 6.25*10$^3$ |
| 4 | 20% | 6.25*10$^3$ |

On the other hand, Hafizi *et al.* [7] demonstrated that, when the base-emitter junction exhibits a graded composition, as it is our case (see Table I), small outdiffusion gives rise to an increase of the DC current gain. The threshold voltage and the DC current gain variations are therefore suitable parameters to detect small Be outdiffusion. Since in our experiments the stress current is pulsed, we expect less pronounced effects than in the DC current stress.

At specific time intervals the stress was interrupted in order to allow the electrical characterization of the sample under investigation. The device characteristics were measured by using a Semiconductor Parameter Analyzer HP4145.

## 4. Experimental results

### *4.1. Variations with the duty cycle*

Figure 2 compares the variations of the threshold voltage versus the effective stress time, i.e. the time during which the devices were on, for stresses 1, 3 and 4 respectively. In the case of stress 4 a considerable increase of the threshold voltage with stress time was observed, while for stresses 1 and 3 the threashold voltage variation is small.

In fact after about 160 minutes of effective stress, the threshold voltage increase for stress 4 was of about 50 mV, while for stress 1 and 3 was lower than 10 mV.

Figure 3 reports the current gain, normalised to the pre-stress value, versus the effective stress time. For stresses 1 and 3 a small increase in the current gain after an effective stress time of 260 minutes can be observed, while for stress 4 the current gain, after a similar increase in the first 20 minutes, begins to reduce quite rapidly.

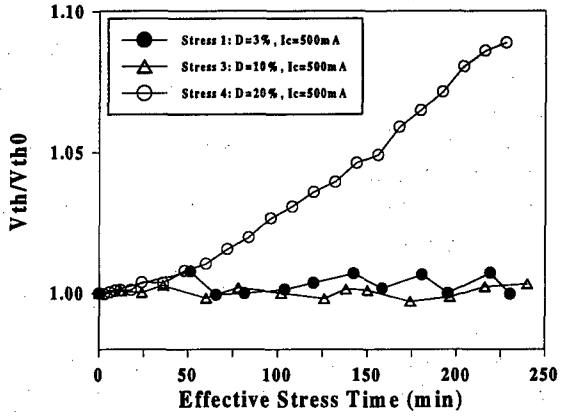



Fig. 2. Ratio of the threshold voltage (Vth) over the initial threshold voltage (Vth0) versus effective stress time.

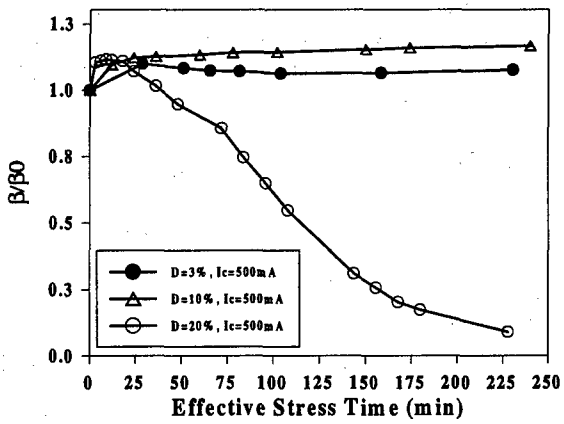



Fig. 3. Ratio of the current gain (β), at $V_{be}$=1.3V, over the initial current gain ($\beta_0$) versus effective stress time.

### *4.2. Variations with the current density*

In Figure 4 the variations of the threshold voltage versus the effective stress time for stresses 1 and 2 is shown. In the case of stress with higher current density a larger increase of the threshold voltage with stress time was observed. Figure 5 depicts the current gain, normalised to the pre-stress value, versus the effective stress time. The current gain

variation is clearly more pronounced in the case of the stress 2.

## 5. Discussion

The difference in the threshold voltage variations observed as a consequence of stresses featuring different duty cycles can be explained in terms of thermal effects. During the stress, the devices were biased in the forward active region and, therefore, they dissipated power and exhibited self-heating. Under dynamic stress condition, as it is our case, the device self-heating, that is the junction temperature, depends also on the thermal constant of the samples [7]. When the duty cycle is short the device on-time is short and therefore the self-heating phenomena are reduced. In Figure 2, the threshold voltage variations induced by the stresses 1 and 3 are practically the same, as pointed out in the previous section. Since the base dopant diffusion is enhanced by the temperature, this result suggests that in both the cases the pulses were not long enough to induces an appreciable junction temperature increase. Indeed, when the duty-cycle was changed to the value of 20%, a pronounced increase of the thresholds voltage was observed.

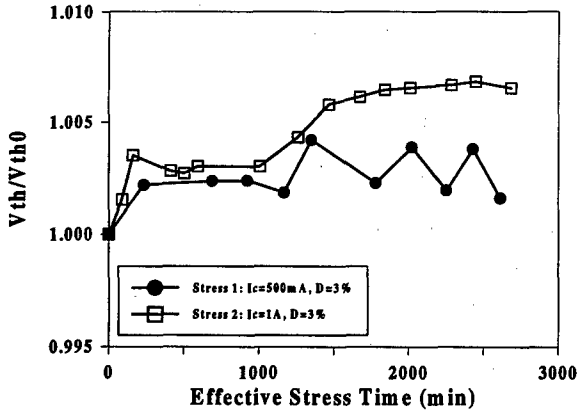



Fig. 4. Ratio of the threshold voltage (Vth) over the initial threshold voltage (Vth0) versus effective stress time.

The behaviours of the threshold voltage is mirrored by the variations of the DC current gain (Figure 3). Still, in this case, no differences between the effects induced by the stress 1 and 3 can be observed, while a clear decrease was shown in the case of stress 4. It is worth noticing that the stress 1 and 3 induced only a slight increase of the DC current gain, also observable in the early phase of the stress 4.

These results demonstrate that the Be outdiffusion is smaller in the case of stress 1 and 3 and larger in the stress 4.

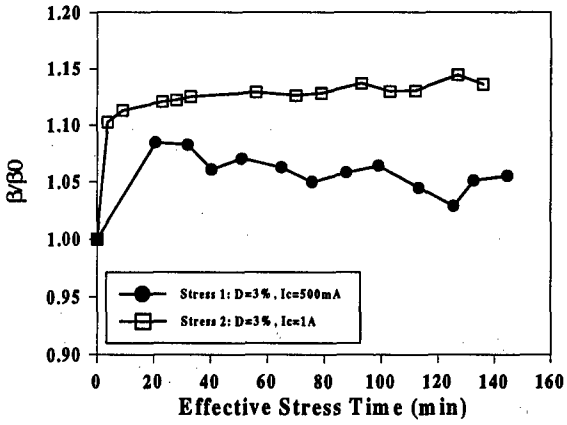



Fig. 5. Ratio of the current gain ($\beta$), at $V_{be}$=1.3V, over the initial current gain ($\beta_0$) versus effective stress time.

The observed differences in the threshold voltage variations when only the stress current was varied can be explained in terms of Recombination Enhanced Impurity Diffusion (REID) mechanism. According to this model during the stress, the energy released by minority carrier recombinations is converted into vibrational energy that results in the annihilation of non-radiative recombination centers and in the generation of defects (probably Ga interstitials) that boosts Be diffusion. Therefore, larger stress current densities give rise to larger injiected minority carriers recombinations and then induce larger degradation, as shown by the Figure 4 and 5. One can observe that the differences in the variations induced by the stress 1 and 2 are a thermal effect, because the dissipated power and therefore the junction temperature increase with changing the stress current from 0.5A to 1A. We can rule out this justification, because no appreciable changes were observed, both on the threshold voltage and on the DC current gain, by comparing the effects induced by the stresses 1 and 3. In our opinion, this means that for duty-cycles less than 10%, the pulses are too short to induce an appreciable device self-heating. This suggests that also when the stress current is as high as 1A an appreciable self-heating for a duty-cycle of 3% has not to be expected.

## 6. Conclusion

In the present paper, we investigated the stability of Be doped AlGaAs/GaAs HBTs under pulsed electrical current stress.

A special circuit, designed and fabricated to this aim, allowed to apply current pulses of various duration (ranging between 3μsec and 20μsec) and various amplitudes (tipically 0.5A and 1A). It has been demonstrated that the pulsed stress technique allows to separate the thermal and REID contribution to the Be diffusion mechanism. This technique could be therefore proposed as an alternative choice to the cooling of the device or to the thinning of the substrate, to investigate the REID contribution to the Be outdiffusion.

From the results obtained at different current and very short duty cycle, we may obtain a further confirmation that a REID mechanism is responsible for HBT degradation at low temperatures.

## Acknowledgements

The authors are grateful to Dr. A.Marty and Dr. J.Tasselli, with LAAS-CNRS (France), for the fabrication of the devices.

## References

[1] Hafizi M., Metzger R., Stanchina W.E., Rensch D.B., Jensen J.F., Hooper W.W., The effects of Base Dopant Diffusion on DC and RF Characteristics of InGaAs/InAlAs Heterojunction Bipolar Transistors, IEEE Electron Device Letters, vol. 13, n. 3, pp.140-142, 1992,

[2] Tanaka S., Shimawaki H., Kasahara K., Honjo K., Characterization of Current-Induced Degradation in Be-doped HBTs Based in GaAs and InP, IEEE Trans. on Electron Devices, vol. 40, n. 7, pp. 1194-1200, 1993

[3] Uematsu M. and Wada K., Recombination-enhanced impurity diffusion in Be-doped GaAs, Applied Physics Letters, vol. 58, n. 18, pp. 2015-2017, 1991.

[4] Kirtania A.K., Das M.B., Chandrasekhar S., Lunardi L.M., Qua G.J., Hamm R.A., Yang L., Measurement and Comparison of 1/f Noise and g-r Noise in Silicon Homojunction and III-V Heterojunction Bipolar Transistors, IEEE Trans. on Electron Devices, vol. 43, n. 5, pp. 784-792

[5] Borgarino M., Voltolini C., Fantini F., Tasselli J., Marty A., Experimental Evaluation of the Minimum Detectable Outdiffusion Lenght for AlGaAs/GaAs HBTs, Solid-State Electronics, vol. 42, n. 3, pp. 325-329, 1998

[6] Dangla J., Dubon-Chevallier C., Filoche M., Azoulay R., Electrical Characterization of the p-type dopant diffusion of highly doped AlGaAs/GaAs Heterojunction Bipolar Transistor grown by MOCVD, Electon. Lett., vol. 26, n. 14, pp. 1061-1063, 1990

[7] Hafizi M., Metzger A., Stanchina W.E., Rensch D.B., Jensen J.F., Hooper W.W., The effects of Base Dopant Diffusion on DC and RF Characteristics of InGaAs/InAlAs Heterojunction Bipolar Transistors, IEEE Electron Device Letters, vol 13, n. 3, pp. 140-142, 1992.

[8] Baibé J.P., Andrieux L., Cazarré A., Camps T., Marty A., Tasselli J., Granier H., Theory and Experiment of the Temperature Dependance of AlGaAs/GaAs HBTs Characteristics for Power Amplifier Application, Solis-State Electronics, vol. 38, n. 2, pp. 279-286, 1995

PERGAMON

Microelectronics Reliability 38 (1998) 1239–1244

MICROELECTRONICS RELIABILITY

# Degradation of performance in MESFETs and HEMTs: simulation and measurement of reliability

Ting Feng, N. Strifas, and A. Christou

*Department of Materials and Nuclear Engineering*
*And Center for Optoelectronic Devices, Interconnects and Packaging,*
*University of Maryland, College Park 20742, USA*

**Abstract**

Low noise GaAs based metal semiconductor field effect transistors (MESFETs) and high electron mobility field effect transistors (HEMTs) are subjected to temperature accelerated testing in order to study the degradation mechanism of the tested devices as a function of temperature and time. The major failure modes are found to consist of a decrease in the zero-bias drain-to-source saturation current ($I_{DSS}$) and pinch-off voltage ($V_p$), and an increase in series resistance of ohmic contact. A degradation model based on the "gate sinking" and degradation of ohmic contact is proposed to simulate the degradation of I-V characteristics. The proposed model includes the effects of a decrease in effective channel thickness, reduction of free carrier mobility, and compensation of free carrier due to the gate metal penetration into the active channel layer, which results in good agreement between the calculated and experimental results.

© 1998 Published by Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Gallium Arsenide based MESFETs and HEMTs have been developed and fabricated for applications in many microwave systems. However, many problems related to reliability still remain. Various life tests have been carried out and several failure modes specific to individual devices have been pointed out. The GaAs MESFET degradation mechanisms reported to date involve gate metallization, gate Schottky contact, and source/drain ohmic contacts, with particular concern about metallizations, surface states, humidity effects, and burn-out[1,2]. However, what has been neglected to date is the extension of device physics models to reliability and the correlation of reliability simulation with device degradation results.

The purpose of this work is to develop a degradation model based on gate metal and ohmic contact degradations. Using this model coupled with SPICE, it is possible to accurately predict lifetimes of devices. To confirm this degradation model, the variations of several dc electrical characteristics of GaAs MESFETs with Ti/Cu gate metallization induced by thermal treatments are investigated. The model is then extended to HEMTs through a charge control approach.

## 2. Reliability testing

GaAs MESFETs with Ti/Cu gate metallization were fabricated with gate width of 200μm and gate length of 1μm. The gate was prepared by electron-beam evaporation of 150nm Ti and subsequent deposition of Cu. The n-type active layer has a typical carrier concentration of $1\times 10^{17}$ $cm^{-3}$ and a thickness of 0.2 - 0.25μm. The ohmic contacts for source and drain were composed of AuGe/Ni/Au alloy with sintering at 410°C to lower contact resistance.

The MESFETs were subjected to temperature accelerated testing at 200°C up to 1000 hours in a vacuum oven, in order to evaluate the thermal stability of the gate structures and device performance degradation with temperature. DC characterization was carried out by HP parameter analyzer, and the Fukui method [3] was used to extract the GaAs MESFETs DC parameters. The drain saturation current $I_{DSS}$, pinch-off voltage $V_p$,

0026-2714/98/$ – see front matter. © 1998 Published by Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00123-1

and source resistance $R_s$ were chosen as the parameters monitoring the degradation of the tested devices. Similarly, the GaAs/GaAlAs based HEMTs were also processed and tested for comparison only.

## 3. Experimental results

Decrease in drain saturation current $I_{DSS}$ and pinch-off voltage $V_p$ and increase of source resistance were measured as a function of time during aging at 200°C, and are shown in figure 1. The decrease of $V_p$ results from the gate metal diffusion into the channel layer, leading to the reduction of effective channel thickness along with the compensation of free carrier by the diffused impurities like Cu, which is called "gate sinking" effect[1, 4]. The decrease of $I_{DSS}$ may be caused by both channel and ohmic contact degradation. The gate sinking was previously reported for Au and Al gate, but not for Cu based metallization. The HEMTs tested at 200°C show a similar degradation. However in the case of HEMTs, the degradation is related to a decrease in the channel charge density due to both band bending and impurity diffusion[1,2].

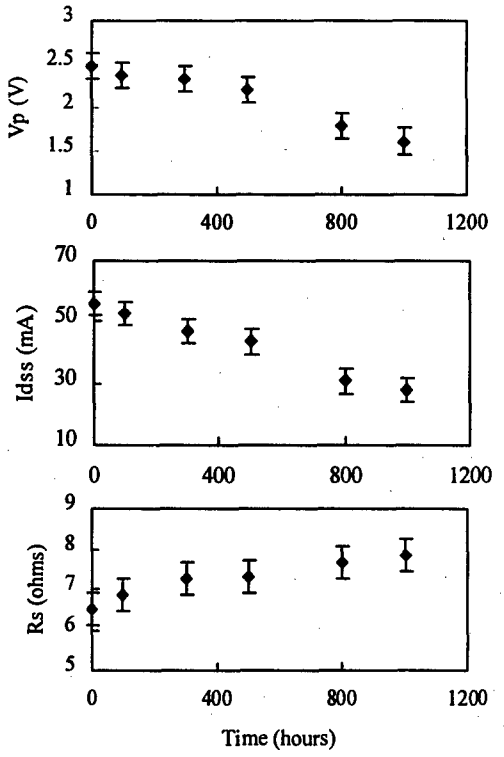



Figure 1. The measured pinch-off voltage $V_p$, drain saturation current Idss, and source resistance as a function of time during aging at 200°C for the tested GaAs MESFETs with Ti/Cu gates.

## 4. Degradation model

A degradation model is proposed based on the observed failure mechanisms related to gate and ohmic contacts. Several factors are taken into account in this proposed degradation simulation model: 1) gate metal diffusion into the channel layer; 2) channel free carrier compensation by the diffused gate metals; 3) decrease of free carrier mobility by the increase of impurities in channel; 4) increase of source resistance by diffusion process. These factors can lead to the degradation of DC parameters such as pinch-off voltage and drain saturation current.

### *4.1 Diffusion of gate metals*

For a single gate contact layer, the diffusion of metal into the channel can be described by the error function:

$$N_{imp}(x) = N_0 erfc(\frac{x}{2\sqrt{Dt}}) \quad (1)$$

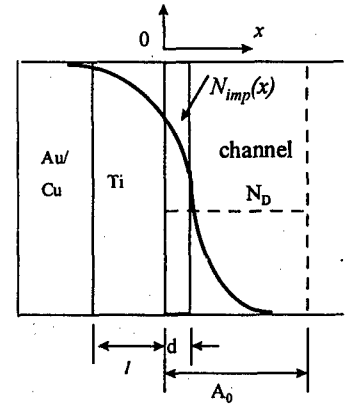



Figure 2. Schematic structure shows the metal impurities diffusion profile into the channel.

However, in most cases the gate of the devices is metallized with two or more metal layers, one is diffusion barrier like Ti, which directly contacts the semiconductor. An overlayer such as Au or Cu is used to lower the gate resistance and increase bonding properties. As shown in figure 2, we assume that the diffusion barrier is stable at relatively low temperatures, and the main diffusive element is the top layer metal such as Au or Cu. By considering of the case of diffusion through a slice into a semi-infinite medium, the analytic solution of concentration profile of diffusive atoms inside the slice and semi-infinite medium can be given by[5]:

$$N1(x) = N_0 \sum_{n=0}^{\infty} \alpha^n \left\{ erfc \frac{(2n+1)l + x}{2\sqrt{D_1 t}} - \alpha erfc \frac{(2n+1)l - x}{2\sqrt{D_1 t}} \right\} \quad (2)$$

$$N2(x) = \frac{2KN_0}{K+1} \sum_{n=0}^{\infty} \alpha^n erfc \frac{(2n+1)l + Kx}{2\sqrt{D_2 t}} \quad (3)$$

where, $N1(x)$ is the concentration profile in the slice (diffusion barrier), and $N2(x)$ is the concentration profile in the semi-infinite medium (channel and substrate), $l$ is the thickness of the slice, $K=(D_1/D_2)^{1/2}$, $\alpha=(1-K)/(1+K)$, $D_1$and $D_2$ are the diffusion coefficients of top layer metal (Au, Cu) in the diffusion barrier layer and GaAs substrate, respectively, $d$ is the thickness of the intermixing layer of gate and channel, which is defined as the diffused metal concentration exceeds solubility of diffused metal atoms in GaAs. The advancement of this intermixing layer is considered to result in the reduction of the effective active channel thickness. The original channel thickness is taken as $A_o$.

### 4.2 Carrier compensation from the diffused metal impurities

The metal impurities like Au or Cu diffused from gate contact can be considered as deep level acceptors, which may compensate the original doping level in the active channel[1,4]. It has been found that Au impurity in GaAs has an acceptor level with energy of 0.09eV above the valence band $E_v$, and Cu impurity has three acceptor levels with first one of 0.14eV above $E_v$[6,7]. Assuming a deep level acceptor concentration $N_{DLA}=N2(x)$, which can be calculated from the above diffusion equations, then the ionized acceptor concentration $N_{DLA}^-$ can be given by[6,7]:

$$N_{DLA}^- = N_{DLA}/[1+4exp((E_{DA}-E_F)/kT)] \quad (4)$$

where $E_{DA}$ and $E_F$ are the positions of deep acceptor energy level and Fermi level, respectively.

The charge neutrality equation can be given as:

$$n + N_{DLA}^- = N_D^+ \quad (5)$$

where $N_D^+$ is ionized donor concentration. For simplicity, the free electron concentration $n$ may be written as:

$$n \approx N_D - N_{DLA}^- \quad (6)$$

which is a function of reliability testing temperature and time.

### 4.3 The decrease of electron mobility due to the increase of impurity concentration

The diffused metals in the channel can increase the impurity scattering of free electrons, which causes the decrease in free electron mobility $\mu$. The free electron mobility is approximately proportional to the inverse of impurity concentration, $\mu \propto 1/N_{imp}$ [7, 8]. Therefore, the decrease of electron mobility $\Delta\mu$, may be simply given in terms of $N_D$ and $N_{DLA}$ as:

$$\Delta\mu/\mu_0 = 1/[1+N_{DLA}/N_D] \quad (7)$$

where $N_{DLA} = N2(x)$, and $\mu_0$ is original free electron mobility in the channel. It is noted that the decrease of free electron mobility is also a function of temperature and time. Therefore, the pinch-off voltage for degraded MESFETs can be simply rewritten as a function of temperature and time according to figure 2:

$$V_p(T,t) = \int_d^{A_o} \frac{q}{\varepsilon} n(T,t,x) x dx \quad (8)$$

### 4.4 Degradation of ohmic contact

The degradation of ohmic contacts is one of main failure mechanisms for long term operation of GaAs devices. The detected increase in contact resistance for AuGe/Ni/Au based ohmic contacts has been attributed to interdiffusion and reaction between AuGeNi alloy and GaAs, which leads to an high resistivity region under contact[1,2]. The change of resistance associated with the interdiffusion process may be described as a semi-empirical equation[9, 10]:

$$\Delta R/R \approx C\,(Dt)^{1/2} \quad (9)$$

where $D$ is the interdiffusion coefficient for the contact and $t$ is heating time, and $C$ is a proportionality constant related to the grain size of alloys. It has been reported that the activation energy of degradation of AuGeNi contact due to interdiffusion effects was found around 0.8-1.0eV[10].

From the above model, the degradation of electrical parameters such as pinch-off voltage, free carrier concentration and mobility, and source resistance may be estimated as a function of testing temperature and time, and by substituting the new data of these parameters into the SPICE program, the degradation of device DC performance can be simulated.

### 4.5 Failure Mechanisms of AlGaAs/GaAs HEMTs

The I-V collapse due to AlGaAs/GaAs interdiffusion and gate degradation has been simulated and failure modes such as kink effect, ohmic contact degradation, and gate failures will be presented.

Under the 200°C accelerated test and dc bias, the hottest spot of the device is between the gate and drain regions. The AlGaAs/GaAs interface degraded via hot electron induced interdiffusion. This will result in two adverse effects: 1) degradation of the 2D electron gas (2DEG) channel resulting in deconfinement of the 2DEG and 2) formation of a potential barrier for current flow through the remaining undegraded 2DEG channel. These two adverse effects will result in a decrease in the drain to source current mainly due to higher resistance of the potential barrier and also due to higher scattering rates. Electron deconfinement in the same region of the device will result in an accompanying decrease in transconductance.

## 5. Simulation results

### 5.1 MESFET simulation

The above degradation model has been applied to simulate the degradation of our tested GaAs MESFET devices with Ti/Cu gates. The data used in the simulation is listed in Table 1.

Table 1 Parameters for Ti/Cu MESFET simulation

| | | |
|---|---|---|
| Cu diffusivity in Ti : $D_1=2.5\times10^{-8}\exp(-2.0eV/kT)m^2/s$ [11] | | |
| Cu diffusivity in GaAs:$D_2=1.0\times10^{-7}\exp(-0.53/kT)m^2/s$ [12] | | |
| Ti thickness : | $l$ | 150nm |
| Channel thickness: | $A_o$ | 250nm |
| Channel donor density: | $N_D$ | $1\times10^{17}$ cm$^{-3}$ |
| Gate length: | $L$ | 1.0 μm |
| Gate width: | $W$ | 250 μm |
| Cu acceptor level in GaAs: | $E_{DA}$ | 0.14eV |

It is noted that since the acceptor level of Cu in GaAs is much larger than kT at room temperature (~0.026eV) and even at 200°C (~0.04eV), the free carrier compensation effect of Cu impurities in the channel may be negligible. However, the solubility of Cu in GaAs has been reported to be approximately $1\times10^{15}$ cm$^{-3}$ for p-type GaAs at a concentration of $10^{17}$ cm$^{-3}$ and approximately $10^{10}$cm$^{-3}$ for intrinsic GaAs, while no data has been reported for n-type GaAs because of the measurement difficulty[12, 13]. When the concentration of the diffused copper at gate and GaAs interface exceeds the solubility of Cu in GaAs, an intermixing compound layer at the interface will grow by interdiffusion or/and reaction. The thickness of the intermixing layer $d$ may be approximately given by[14]:

$$d^n = k_g(t-t_0) \tag{10}$$

where $n$ =1 for reaction-control growth, and $n$=2 for diffusion-control growth, $t_0$ is the starting growth time during aging when the solubility is reached, and $k_g$ is proportionality factor.

Assuming the solubility of Cu in n-type GaAs to be approximately $5\times10^{15}$cm$^{-3}$, the degradation of the pinch-off voltage $V_p$ is simulated as shown in figure 3, in good agreement with the measured data. The simulated contact resistance is also shown in figure 4 and compared with the measured data.

In addition, by substituting the variations of pinch-off voltage, effective thickness, net free carrier density, electron mobility, and source resistance with time into the SPICE program, the drain saturation current is simulated as a function of aging time as shown in figure 5.

Figure 6 shows the evidence of compound formation at the gate and channel interface of the tested MESFETs after aging at 200°C for 1000 hours obtained from the back-etching of the GaAs substrate.

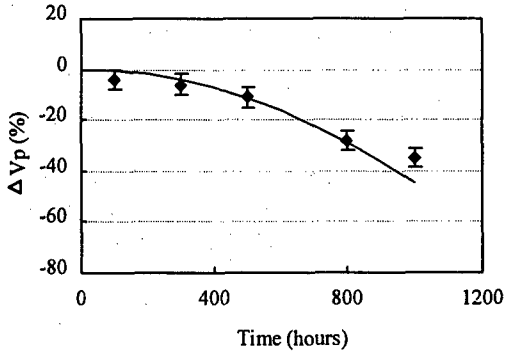



Figure 3. The calculated degradation of pinch-off voltage as a function of aging time for Ti/Cu MESFETs.

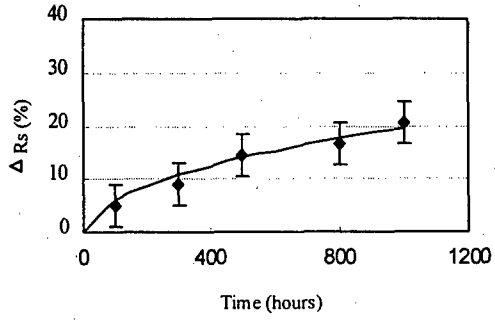



Figure 4. The calculated degradation of source resistance as a function of aging time for Ti/Cu MESFETs.

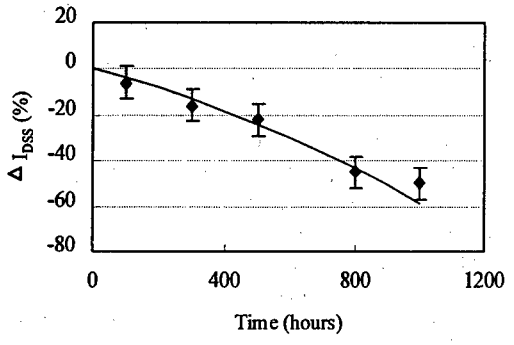



Figure 5. The calculated degradation of drain saturation current as a function of aging time from SPICE based on the input data calculated from the degradation model for Ti/Cu MESFETs

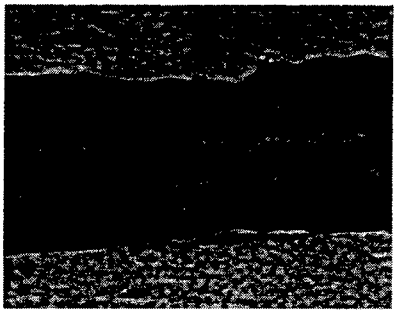


Figure 6. SEM image of gate interface obtained by back-etching away GaAs substrate shows the evidence of compound formation for Cu/Ti gated MESFETs after aging at 200°C up to 1000 hours.

### *5.2 HEMT simulation*

The interdiffusion between AlGaAs and GaAs decreases the donor layer thickness resulting in a current-voltage collapse. The I-V changes have been simulated using the parameters in Table 2.

Figure 7 shows the I-V collapse simulated for different temperatures as a result of AlGaAs/GaAs interdiffusion. The drain to source current decrease can be explained from the decrease of the channel mobility and free electron concentration. The decrease of the 2DEG channel thickness would also limit the drain to source current. Similarly, the gate-metal/AlGaAs interdiffusion has been simulated and coupled with experimental results.

By virtue of the large transient electron current density toward the channel through the gate, gate-metal/AlGaAs interdiffusion occurs. The penetration of the gate metal within the channel and the formation of the potential barrier for current flow under the gate edge will result in a decrease of the pinch-off voltage. The kink effect may also be simulated by introducing deep donor levels where emission time decreases as the electric filed increases. Hence when the drain voltage is applied, the electrons of the 2DEG gain energy from the channel and hot electrons are injected into the AlGaAs layers where they are captured by the deep levels.

Table 2 Parameters for HEMTs simulation

| | | |
|---|---|---|
| Gate to source voltage | $V_{gs}$ | 0.6 V |
| Threshold voltage | $V_{th}$ | 0.1 V |
| Gate length | $L_g$ | 1 μm |
| Gate width | W | 100 μm |
| Resistance temperature coeff. | α | $-4.2X10^{-3}$ $K^{-1}$ |
| AlGaAs layer thickness | $d_i$ | 400 Å |
| Dielectric permittivity of AlGaAs | | $1.08X10^{-10}$ F/m |
| Effective thickness of 2DEG | t | 40 Å |
| Total $n_s$ doping concentration in channel | $N_T$ | $2.6X10^{15}$ $cm^{-3}$ |

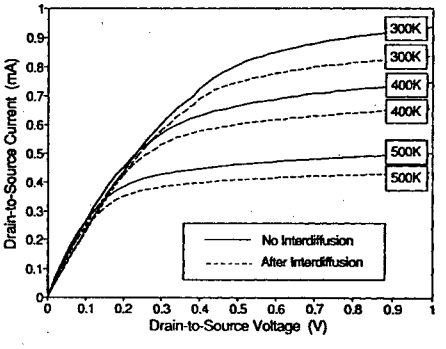



Figure 7. The I-V characteristics collapse due to AlGaAs/GaAs interdiffusion.

## 6. Conclusions

We have successfully simulated through the developed device physics reliability models failure mechanisms pertaining to MESFETs and HEMTs. This represents the first successful correlation of the device experimental results. MESFET degradation mechanisms centeres around metallization degradation while HEMT degradation mechanisms includes 2DEG deconfinement, the kink effect in addition to gate-AlGaAs interdiffusion.

**Acknowledgement:**

The investigations were supported by the Office of Naval Research, DARPA and partially by ARL through the Microelectronics Research Collaboration Program.

**References:**

[1] A. Christou, Reliability of Gallium Arsenide MMICs, Wiley, Chichester, 1992.
[2] E. Pollino, Microelectronic Reliability Vol II: Integrity Assessment and Assurance, Artech House, Norwood, 1989.
[3] H. Fukui, Determination of the basic device parameters of a GaAs MESFET, *Bell Syst. Tech. J.*, **58**, pp771 (1979).
[4] J.C. Irvin, *GaAs FET Principles and Technology* (Eds. J. W. DiLorenzo and D. D. Khandelwal), ch. 6, Artech House, Dedham (1982).
[5] Crank, The Mathematics of diffusion, Clarendon Press, Oxford, 1975.
[6] Sze, Physics of semiconductor devices, 2nd, Wiley, 1981.
[7] A.G. Milnes, Deep Impurities in Semiconductors, Wiley, New York, 1973.
[8] M. Shur, GaAs devices and circuits, Plenum Press, New York, 1986.
[9] J. Kucera, B. Million, J. Zidu, V. Hermasky, *Thin solid Films*, **230** (1993) 183.
[10] K. Mizuishi, H. Kurono, H. Sato, and H. Kodera, *IEEE Trans. on Electron Devices* **ED-26** (1979), 1008.
[11] O. Taguchi and Y. Iijima, *Philosophical magazine A*, **72** (1995), 1649.
[12] Shaw, Atomic diffusion in semiconductors, Plenum Press, London, 1973.
[13] R.N. Hall and J.H. Racette, *J. Appl. Phys.* **35** (1964), 379.
[14] K.N. Tu, J.W. Mayer, and L.C. Feldman, Electronic thin film science, Maxwell Macmillan, New York, 1992.

PERGAMON

Microelectronics Reliability 38 (1998) 1245–1250

MICROELECTRONICS RELIABILITY

# A new method for temperature mapping on GaAs field effect transistors

E. Martin[a], J. P. Landesman[a], P. Braun[b], A. Fily[a]

[a]*Thomson-CSF, Laboratoire Central de Recherches, Domaine de Corbeville, 91404 ORSAY Cedex, France*
[b]*United Monolithic Semiconductors GmbH, Wilhelm-Runge-Str. 11, 89081 Ulm, Germany*

**Abstract**

In this paper, a method is described, how to apply the technique of spatially resolved photoluminescence (PL) spectroscopy for the measurement of the local channel temperature in GaAs-based field effect transistors. This spectroscopic technique uses a focused laser beam which scans directly the surface of a chip inside its package. The temperature is deduced from the corresponding wavelength shift of the PL peak. In the case of a typical heterostructure-based transistor (like the pseudomorphic high electron mobility transistors studied here) a spatial resolution of 1 μm and a temperature resolution of ±1°C is demonstrated.
© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Local channel heating in field effect transistors (FETs) limits the output power of discrete devices or monolithic microwave integrated circuits (MMICs) designed for high power applications, especially when GaAs is used [1], and also more generally limits the lifetime of any type of FET due to thermally activated degradation or failure mechanisms [2]. Temperature mapping is presently done with either of the two widely used commercial techniques: infrared microscopy [3] or liquid crystal thermography [4]. Infrared microscopy is limited mainly in terms of spatial resolution (at best a few micrometers) which is not compatible with technologies with submicron gate length. On the other hand, the liquid crystal thermography technique was initially designed to visualize hot spots [5], but quantitative evaluation of the local temperature using this technique is not straightforward, mainly due to difficulties in controlling the coating of the chip by the liquid crystal and in stabilizing the temperature of the hot plate with the required accuracy.

In this paper, an approach is proposed based on photoluminescence (PL) spectroscopy [6, 7] for a non destructive mapping of local channel temperatures in GaAs FETs with submicron gate length.

The use of spatially resolved PL spectroscopy to evaluate temperature distributions in GaAs devices has already been demonstrated, e. g. for the problem of juntion heating in GaAs laser diodes [8].

## 2. Experiment

The devices that were investigated are pseudomorphic high electron mobility transistors (PHEMTs) fabricated at United Monolithic Semiconductors. The semiconductor material is grown by

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00095-X

molecular beam epitaxy on a semi-insulating GaAs substrate (typical layer sequence given in table 1).

Table 1.
Typical layer sequence for the PHEMTs.

| Layer | Thickness (nm) | Doping level |
|---|---|---|
| GaAs cap | 50 | $\geq 10^{18}$ cm$^{-3}$ (Si) |
| $Ga_{1-x}Al_xAs$ barrier | 30 | no doping |
| Delta doping | | $\geq 10^{12}$ cm$^{-2}$ (Si) |
| $Ga_{1-x}Al_xAs$ spacer | 3 | no doping |
| InGaAs channel | 12 | no doping |
| $GaAs/Ga_{1-x}Al_xAs$ | 500 | no doping |
| Substrate | | semi-insulating |

Figure 1 shows a cross section of the source-drain region of the transistor, with the T-shape gate technology. This PHEMT technology is used for low noise applications up to 95 Ghz as well as for medium power applications up to 60 Ghz [9], depending on the gate length.

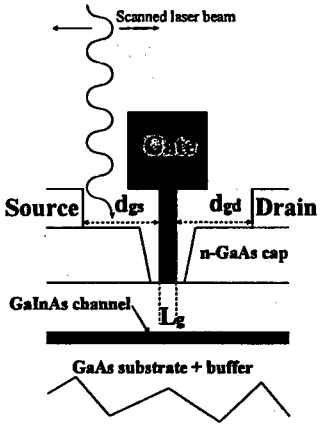



Fig. 1. Cross section of a transistor. The most important geometrical data are: source-drain distance: 2.5 μm; $L_g$=0.25 or 0.15 μm.

For this specific study, two kinds of devices were investigated. First, a 4-gate transistor, with a gate length of 0.25 μm and a gate width of 50 μm (overall gate width: 200 μm) was systematically measured for a number of different operating points, under DC bias. Figure 2a shows the I-V characteristics for this device, which is representative of the technology for medium power applications. Second, a 2-gate transistor with a gate length of 0.15 μm (gate width 75 μm), selected from the low noise devices, was measured at a single DC bias. Figure 2b shows the I-V characteristics of this second transistor.

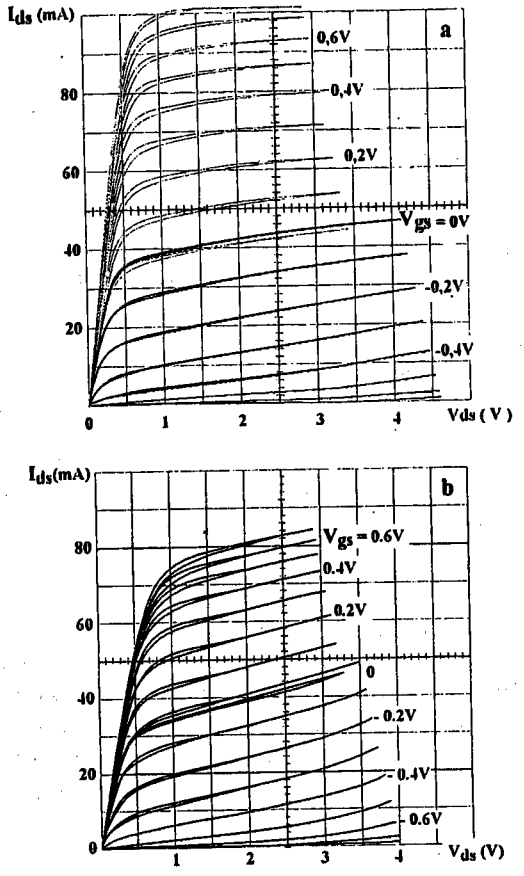



Fig. 2. I-V characteristics of the transistors: a) 4-gate structure, $L_g$=0.25 μm; b) 2-gate structure, $L_g$=0.15 μm.

The excitation source was the 647 nm line of a $Kr^+$ laser, with a power density ≈10 kW/cm$^2$.

Focusing of the spot at the diffraction limit onto the sample inside its microwave package (BMH60) was made through a long working distance objective in an optical microscope (50X). The PL spectrum produced is back-transmitted through the microscope objective and analyzed in a grating spectrometer equipped with a set of confocal holes to enhance the spatial resolution. Using a high precision x-y stage the samples are positionned with a reproducibility of ≈±0.1 μm.

A quantitative determination of the spot size under the conditions used for this study (in particu-

lar, confocal hole diameter =100 μm, corresponding to a virtual aperture =2 μm at the sample surface) was done in a separate experiment. This involved scanning over the cross section of a reference heterostructure material, including two layers with distinct spectral PL features. The spot size thus measured is close to 0.7 μm.

## 3. Results and discussion

Figure 3 shows the PL spectrum obtained when the laser beam is focused onto the channel region.

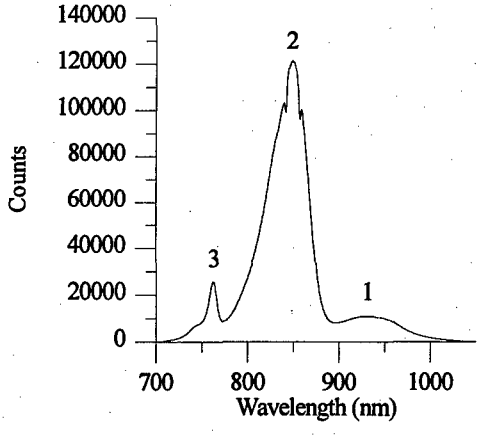



Fig. 3. PL spectrum mesured in the channel region.

This spectrum, recorded with a low spectral resolution (150 grooves/mm grating) shows three main peaks: i) the broad structure labelled 1 (920-950 nm) corresponds to transitions occurring in the GaInAs channel; ii) the peak labelled 2 (865 nm) corresponds to the GaAs cap layer; iii) the sharp structure labelled 3 (820 nm) is induced in the GaAs/GaAlAs superlattice buffer (SB) incorporated close to the GaInAs channel for technological reasons (improvement of the epitaxy and barrier to parasitic conduction).

Upon heating, the band associated with the $Ga_{1-x}In_xAs$ channel does not show a simple spectral shift, but also changes in shape. This is due to the high electron density in this layer [10]. On the other hand, the temperature-induced modification of the SB peak consists mainly in a clear redshift associated with heating, as shown in figure 4. The set of curves in figure 4 was obtained by heating the unbiased device to a controlled temperature (in the range 20-100°C), using a higher resolution grating (600 grooves/mm) to measure the SB peak. As can be seen in the inset of figure 4, the change in wavelength is highly linear in the range 20-100°C. The accuracy on the temperature determination derived from this procedure is $\frac{\delta(\Delta T)}{\Delta T} = \pm 0.9\%$ leading to an error smaller than 1°C up to 100°C. This calibration procedure was repeated for all the transistors which were investigated here (indeed, the slope of the wavelength-temperature curve was found dependent upon the specific starting wafer used for the process of the transistor).

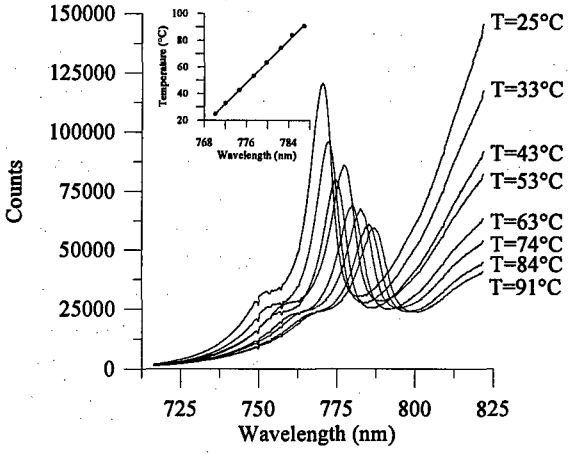



Fig. 4. Temperature dependence of the SB peak (high resolution grating). The inset shows the calibration curve.

In a second step PL measurements were made with the laser beam focused on the transistor DC biased with different operating points. Figure 5 shows the spectra recorded in the drain-gate region.

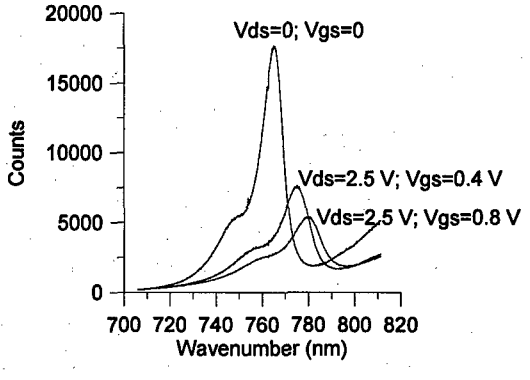



Fig. 5. Spectra recorded in the drain side region with the transistor biased under three different conditions ($L_g$=0.25 μm).

As can be seen, when the drain-source and/or the gate-source voltage increases, the peak shifts indicating heating.

There are two reasons responsible for this effect. The first one is direct heating due to absorption of the laser beam in the different layers of the structure. This effect was quantified by measuring the heating produced on an un-biased transistor by the $Kr^+$ laser beam with different power densities. It was observed that in the range 20-100°C, this additional heating is at most ≈0.1°C for the conditions of the present measurements. The second perturbation is due to the photocurrent generated during laser illumination. It is observed as a slight increase of the drain-source current $I_{ds}$. The magnitude of this effect is of the order of a few % for the conditions used in this series of experiments. By calculating roughly the total number of electron-hole pairs that can be generated in the whole structure (described in table 1) and transferred in the $Ga_{1-x}In_xAs$ channel, this photocurrent is estimated of the order of a few μA/mm. The major part of the photocurrent generation occurs in the GaAs substrate, as shown by the fact that the absolute value of the perturbation measured on $I_{ds}$ is independent on the gate voltage $V_{gs}$ (and in particular is the same for positive and negative values of $V_{gs}$). The overall perturbation (direct heating + photocurrent) was studied as a function of the laser beam power density on a biased transistor, and specified to ≈0.5°C for the conditions used here.

### *3.1. Temperature linescans from source to drain, across the gate*

To assess the spatial resolution of the method, the laser beam was scanned across the channel region, from the source to the drain metallizations, passing over the gate electrode. This was done on the two types of transistors, DC-biased at a single operating point ($V_{ds}$=2.5 V / $V_{gs}$=0 for the 0.25 μm one, $V_{ds}$=1.5 V / $V_{gs}$=0 for the 0.15 μm one). The two curves in figure 6 indicate that in both cases we are able to observe an asymmetry in the temperature distribution (higher on the drain side of the gate). This effect is of course expected for the FET structures [11, 12]. The observation, together with the measurement of the beam diameter, confirms that the spatial resolution in our thechnique is better than 1 μm.

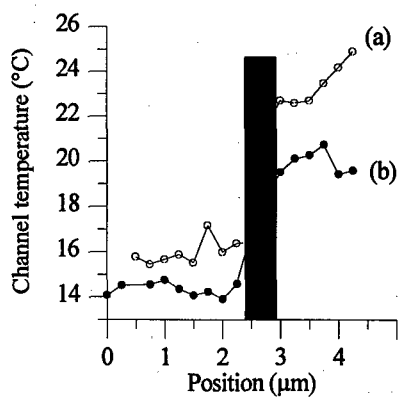



Fig. 6. Local heating measured across a 0.15 μm transistor ($V_{ds}$=1.5 V; $V_{gs}$=0) (a) and 0.25 μm transistor ($V_{ds}$=2.5 V; $V_{gs}$=0) (b). The black zone represents the gate metallization. The channel temperature is referred to the temperature (measured with a termocouple) at the backside of the chip.

### *3.2. Temperature linescans along the gates*

The transistors were measured by scanning the laser beam along each of the gates, on the drain side.

For the 0.25 μm gate length transistor, this was done systematically as a function of $V_{ds}$ and $V_{gs}$ on all 4 gates. The corresponding temperature linescans are shown in figures 7 and 8. The temperature difference between the channel and the chip package backside (which was seen to increase by less than 1°C for all the measurements) was plotted.

For the $V_{gs}$=0 series (figure 7) very little dispersion between the four gates is observed, except for the fact that gate n° 4 (which is on the side of the transistor) is slightly colder than the other three. Different possible origins are investigated at the moment to explain this effect (detailed geometry of the gate, that is realized by electron beam lithography, or of the recess,...).

From the series of curves on figure 8 (different values of $V_{gs}$), the same trend as in figure 7 is seem. However, an additional point is observed: for the two highest powers ($V_{gs}$=0.6 and 0.8 V) the temperature saturates at one end of gate n° 3. These operating points were the last ones investigated. This observation seems to indicate the onset of a degradation mechanism occuring on gate n° 3. The I-V characteristics of these transistors were checked after the series of PL measurements, and a clear degradation of the transistor was confirmed: $I_{ds}$ was

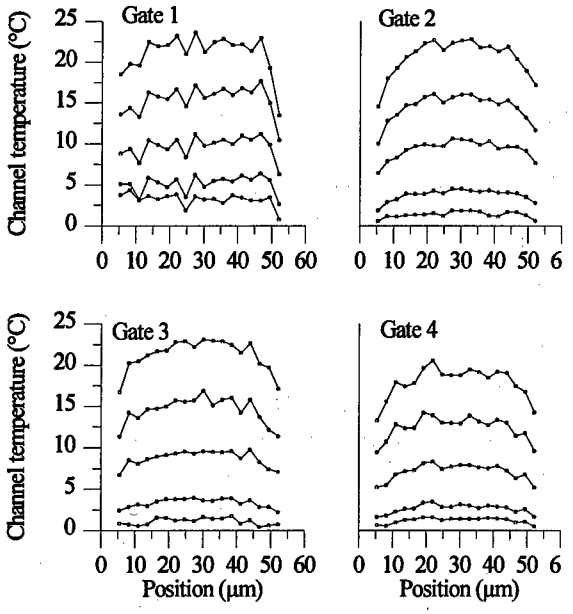



Fig. 7. Local temperature linescans measured in the 0.25 μm transistor as a function of $V_{ds}$ ($V_{ds}$=0.5, 1, 1.5, 2 and 2.5 V); $V_{gs}$=0.

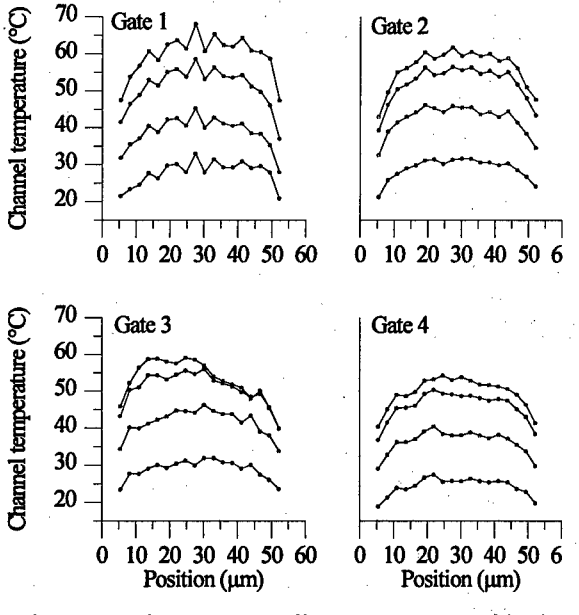



Fig. 8. Local temperature linescans measured in the 0.25 μm transistor as a function of $V_{gs}$ ($V_{gs}$=0.2, 0.4, 0.6 and 0.8 V); $V_{ds}$=2.5 V.

found systematically reduced by 25% as compared to the initial state of the transistor. Therefore, a localized degradation on gate n° 3 can be assumed. This degradation is probably due to the unshielded electrical environment (especially with respect to ESD) of the PL setup used for this analysis.

For the 0.15 μm gate length transistor a temperature scan along the gate was performed. The result of this scan is shown in figure 9.

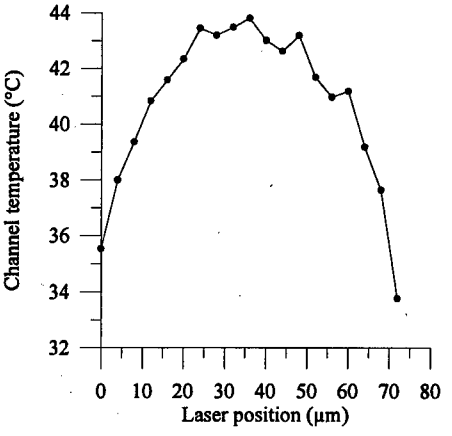



Fig. 9. Local temperature linescan measured along the drain side of the 0.15 μm transistor ($V_{ds}$=1.5 V; $V_{gs}$=0.8 V).

### *3.3. Discussion. Comparison with thermal resistance calculations*

Figure 10 shows, for the 0.25 μm gate length transistors, the evolution of ΔT (temperature difference between the channel and the back side) as a function of the total power dissipated. ΔT is obtained by averaging all the measurements on each linescan of figure 7 and 8, and the power is obtained as $V_{ds}$ x $I_{ds}$ (it was checked that the dissipation in the gates can be neglected).

The first observation is that the measured temperature is significantly lower (at least for the 0.25 μm device) than the value predicted from a « classical » thermal resistance model [13]. It should be noted that for the calculated lines on figure 10 a constant value for the thermal conductivity for the GaAs substrate was used (taken as the room temperature value 0.47 $Wcm^{-1}K^{-1}$). In addition only the thermal resistance of the chip itself has been taken into account. All the other layers in the package were ignored.

Another point that can be seen is the apparent change in dissipation area when $V_{gs}$ is turned on with a positive value. A step on the temperature curves on figure 10 (for the 0.25 μm device) is

identified between 0.6 and 0.7 W/mm, corresponding to the change between the experiments with $V_{gs}$=0 and those with $V_{gs}$>0. This behaviour was confirmed by measuring the channel temperatures produced for two bias conditions with a constant $I_{ds}$ x $V_{ds}$ value, one with $V_{gs}$<0 and the other with $V_{gs}$>0: the point with $V_{gs}$>0 lead to a slighly higher channel temperature. These investigations indicate that the channel temperature is not just a function of the total dissipated power, but also depends on $V_{gs}$. These details can be experimentally observed with the method proposed here.

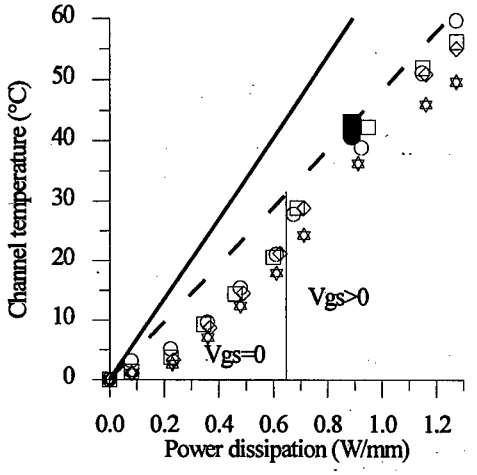



Fig. 10. Average temperature increase for the 0.25 μm transistor (○gate 1, □gate 2, ◊ gate 3 and ✡ gate 4). The single point investigated for the 0.15 μm is also shown (●gate 1 and ■ gate 2). The straight lines are calculated using Cooke's model for the thermal resistance [13] (full line: $L_g$=0.25 μm; dashed line: $L_g$=0.15 μm).

## 4. Conclusions

It was shown that our approach for the mapping of local channel temperatures using spatially resolved PL spectroscopy gives a spatial resolution of the order of 1 μm and a temperature resolution of ±1°C for the PHEMT structure investigated. This was possible due to the specific GaAs epitaxial layer structure used for those devices. Using such a measurement method, in conjunction with an accurate thermal simulator, will lead to better understanding for the details of power dissipation in the transistors.

## Acknowledgements

The authors thank Didier Adam, Frédéric Diana and Françoise Lemaire for help with the experiments and fruitful discussions. This work was partially funded by the European Community under the SMT Programme « NOSTRADAMUS » (contract SMT-CT95-2024).

## References

[1] Wemple S. H., Huang H. C.; *« GaAs FET principles and Technology »*, edited by J. V. Di Lorenzo and D. D. Khandelwal (Artech House, Dedham, Massachusets, 1982), pp. 309-347.

[2] Magistrali F., Tedesco C., Zanoni E., Canali C.; *« Reliability of Gallium Arsenide MMICs »*, edited by A. Christou (John Wiley, Chichester, UK, 1992), pp. 101-189.

[3] Canali C., Chiussi F., Donzelli G., Magistrali F., Zanoni E.; Microelectron. Reliab. **29**, 117 (1989).

[4] Stephens C. E., Sinnadurai F. N., J. Phys. E: Scientific Instruments **7**, 641 (1974).

[5] Hirayama N., Nikawa K., Nakagiri M.; Proceedings ISTFA 1986, pp. 139-144.

[6] Perkowitz S., *« Optical Characterization of Semiconductors; Infrared, Raman and photoluminescence Spectroscopy »*, Academic Press 1993.

[7] Landesman J. P., Depret B., Fily A., Nagle J., Braun P.; Appl. Phys. Lett. **72** (11), 1338 (1998).

[8] Hall D. C., Goldberg L., Mehuys D.; Appl. Phys. Lett. **61**, 384 (1992).

[9] *« Microwave Products Catalogue »*, Section 7, UMS (1997).

[10] Gilpérez J. M., Sanchez-Rojas J. L., Muñoz E., Calleja E., David J. P. R., Reddy M., Hill G., Sanchez-Dehesa J.; J. Appl. Phys. **76** (10), 5931-5944 (1994).

[11] Ghione G., Naldi C. U.; Proceedings International Electron Devices Meeting (1989), p. 147.

[12] Ostermeir R., Brunner K., Abstreiter G, Weber W.; IEEE Trans. Electron. Devices **39** n° 4, 858-863 (1992).

[13] Cooke H. F.; Microwaves RF **25**, pp.85-87 (1986).

PERGAMON

Microelectronics Reliability 38 (1998) 1251–1260

MICROELECTRONICS RELIABILITY

# Reliability of industrial packaging for microsystems

R. de Reus [a,1], C. Christensen [a], S. Weichel [a], S. Bouwstra [a], J. Janting [b], G. Friis Eriksen [b], K. Dyrbye [b], T. Romedahl Brown [b], J.P. Krog [c], O. Søndergård Jensen [c], and P. Gravesen [c]

[a] *Mikroelektronik Centret, DTU bldg. 345-east, DK-2800 Lyngby, Denmark*

[b] *Grundfos A/S, DK-8850 Bjerringbro, Denmark*

[c] *Danfoss A/S, DK-6430 Nordborg, Denmark*

**Abstract**

Packaging concepts for silicon-based micromachined sensors exposed to harsh environments are explored. By exposing the sensors directly to the media and applying protection at the wafer level the packaging and assembly will be simplified as compared to conventional methods of fabrication.

Protective coatings of amorphous silicon carbide and tantalum oxide are suitable candidates with etch rates below 0.1 Å/h in aqueous solutions with pH 11 at temperatures up to 140 °C. Si–Ta–N films exhibit etch rates around 1 Å/h. Parylene C coatings did not etch but peeled off after extended exposure times at elevated temperatures. The best diamond-like carbon films we tested did not etch, but delaminated due to local penetration of the etchants.

Several glue types were investigated for chip mounting of the sensors. Hard epoxies, such as Epo-tek H77, on the one hand exhibit high bond strength and least degradation and leakage, but on the other hand introduce large sensor output drift with temperature changes. Softening of the Epo-tek H77 was observed at 70 °C.

An industrially attractive thin-film anodic silicon-to-silicon wafer bonding process was developed. Glass layers are deposited at 20 nm/s (1.2 μm/min) by electron-beam evaporation and bond strengths in excess of 25 N/mm$^2$ are obtained for bonding temperatures higher than 300 °C.

Through-hole electrical feedthroughs with a minimum line width of 20 μm and a density of 250 wires per cm were obtained by applying electro-depositable photo-resist. Hermetically sealed feedthroughs were obtained using glass frits, which withstand pressures of 4000 bar. © 1998 Elsevier Science Ltd. All rights reserved.

## 1 Introduction

Due to the large diversity in micromachined sensor applications, to date no generic packaging concepts exist. In many occasions chip packaging is not an integral part of the chip design. With a few exceptions, packaging is bulky and performed at chip level, usually resulting in complicated and labor-intensive assembly [1].

Conventional packaging of silicon sensors for aggressive media applications involve stainless steel housings, silicone oil filling, and steel membrane seals to avoid direct exposure of the sensor to the environment. By exposing the sensors directly to the media the packaging volume and assembly costs may be reduced considerably. Such an approach requires adapted sensor designs, wafer-level protective coatings, and reliable chip mount techniques and materials.

This paper reviews the work of an industrial collaboration toward reliable packaging concepts for silicon sensors exposed to harsh environments.

[1] e-mail: reus@mic.dtu.dk

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00149-8

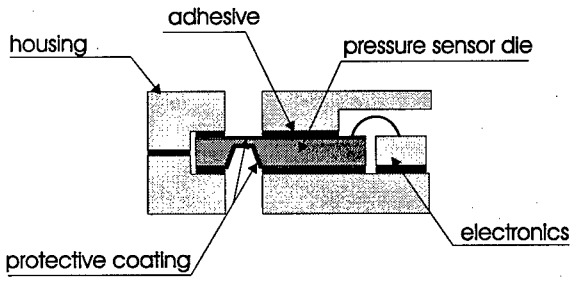



Fig. 1. Cross-sectional drawing of a differential pressure sensor directly exposed to the media. The sensor chip is coated at both sides and mounted in a housing using adhesives.

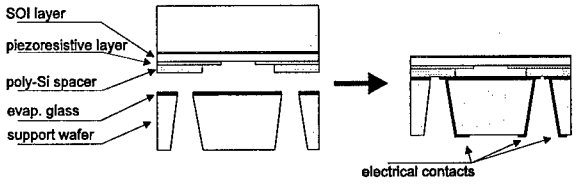



Fig. 2. Flat-surface absolute pressure sensor concept. Poly-silicon piezoresistors and spacers are deposited on an SOI-wafer and anodically bonded to a support wafer. After KOH-etching the SOI device layer forms the membrane and a flat surface is obtained. Electrical contacts are accessible through the holes in the support wafer.

## 2 Sensor and packaging concepts

Several approaches to packaging of silicon sensors for harsh environments are under investigation. The most straightforward method is shown in Fig. 1. Based on a conventional piezoresistive differential pressure sensor, only small alterations to the pressure sensor design itself are required. The bond pads are all positioned at one end of the sensor die. This creates free surface area for adhesive mounting of the chip in a metal housing and facilitates hybrid integration with integrated circuitry. The sensor is exposed directly to the environment via holes in the housing and must therefore be protected with a coating at both sides prior to chip mounting. Packaging induced stress, penetration of the media through the adhesive and the lack of availability of a suitable coating material are key obstacles for a successful industrialisation of this concept.

Because poor step coverage of the thin (less than 1 $\mu$m) protective coating turns out to be a critical issue, the second concept aims at creating a flat-surface sensor. We reported on such a pressure sensor type earlier [2]. Here we present a new

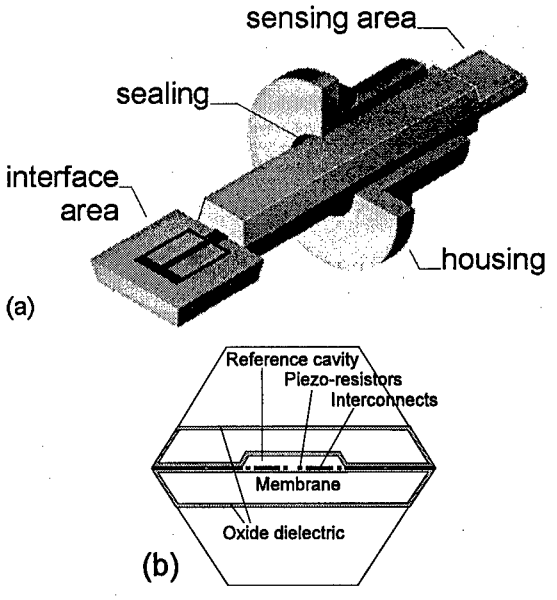



Fig. 3. Drawing of the 'needle sensor' concept (a). The interface area is separated from the medium by a hermetic sealing to the housing. The width of the sensor is typically 1 mm. (b) Cross-sectional drawing of the sensing area showing the bonded wafers, membrane, reference cavity, and integrated interconnects.

type of absolute pressure sensor (single-sided, see Fig. 2). A piezoresistive layer and poly-Si spacers are deposited onto an SOI-wafer and patterned. The spacer layer contains electrical interconnects and is anodically bonded to a support wafer. Access to the poly-Si contacts is provided via contact holes in the support wafer. The substrate-Si of the SOI wafer is then removed by wet chemical etching, leaving the SOI device layer as membrane. The buried oxide serves as etch stop layer and yields a smooth surface suitable for protective coatings which otherwise would suffer poor step coverage. Further advantages of this sensor design are that the active sensing areas are at the backside of the membrane and that electrical contacts are provided at the backside of the sensor. To minimize the chip area and to maintain a rigid support a single contact hole with multiple electrical feedthroughs is preferred. A further challenge is the wafer bonding process to poly-Si. An elaborate paper on the fabrication and characterization of this sensor will be published elsewhere [3].

The innovative 'needle sensor' concept is presented in Fig. 3. This very small sensor (lateral dimensions of the sensor head are $1 \times 1$ mm$^2$) integrates the packaging concept and electrical

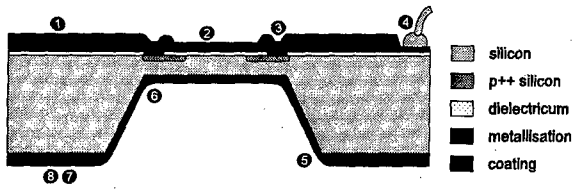



Fig. 4. Requirements for protective coatings. A cross section of a typical piezoresistive pressure sensor is shown. Several critical properties for the coating are identified. Refer to the text for an explanation of the running numbers.

feedthroughs in the device. As in the concept mentioned above, the electrical active parts of the sensor are 'inside' the sensor at the backside of the membrane. Specific challenges to produce this sensor are wafer bonding, a conformal sensor coating of the three-dimensional structure and hermetic sealing between sensor and housing.

In the following section we report on the materials and processes required to accomplish the packaging concepts mentioned above.

## 3 Materials

### 3.1 *Protective coatings*

Protective coatings applied at wafer-level reduce requirements on packaging and assembly. Applying protective coatings as a solution to the sensor concept shown in Fig. 1 requires a number of properties for the coating. In Fig. 4 a number of requirements is identified:

(1) Corrosion resistance: the maximum allowable thickness of the coating and minimum required lifetime sets the upper limit of the etch rate in the media of interest.
(2) Low residual stress and small thickness: to limit the reduction of sensitivity due to stiffness changes in the membrane. Also, a high tensile stress may result in cracking of the coating. Furthermore, the adhesion of the coating, sensor performance, and electrical properties of interconnects may be affected by high stress levels.
(3) Step coverage: poor coverage over interconnects and contact windows are sites where degradation of the sensor will initiate.
(4) Patternable: in many cases it is desired to pattern the protective coating for access to bond pads. Patterning in a batch process, such as wet etching, is preferred.
(5) Coverage of sharp corners: a conformal coating is required.
(6) Coverage of deep cavities: a conformal coating is required down to the bottom of the cavity.
(7) Pinhole density: usually no pinholes are allowed in the exposed area of the sensor. Etchants will penetrate the coating and degrade electrically active components or underetch, eventually resulting in an undesired lift-off of the coating. In case the pinholes are due to particulate contamination, the pinholes may be eliminated by growing thicker films.
(8) Electrical properties: a dielectric film is required to insulate electrical components on the sensor from electrically conducting media.

Furthermore, good adhesion and good diffusion barrier properties are desired. Although the above requirements all are essential, corrosion resistance (1) and low pinhole density (7) may be most important, since these properties cannot be circumvented by alternate sensor designs or materials combinations.

Protection against acidic environments usually is not a problem and conventional materials from semiconductor industry can be used to encapsulate, a.o., metal lines on silicon substrates, which do not etch in, e.g. hydrofluoric acid [4].

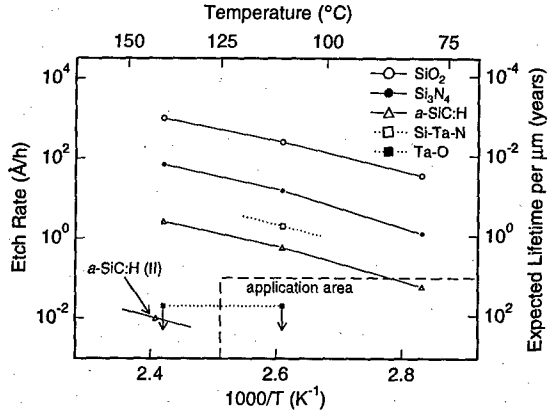



Fig. 5. Etch rates of several coating materials in alkaline solutions. Data points connected with solid line pH 11, dotted line pH 13. The application window is marked. For Ta–O only an upper limit of the etch rate is given (indicated by arrows) since no detectable loss occurred after 3 months of exposure.

Alkaline environments form a greater challenge. The specifications for our sensor applications are a minimum lifetime of ten years for maximum film thickness of one micron in alkaline solutions with pH 11 and temperatures up to 125 °C. Other media of interest include refrigerants, lubricants, and hydraulic oils containing additives. The application window is indicated in Fig. 5, which also shows that conventional passivation layers such as silicon dioxide ($SiO_2$) and silicon nitride ($Si_3N_4$) do not fulfill our requirements (this work and [4]).

Grain boundaries are expected to be weak points for corrosion resistance, very similar to diffusion barrier performance [5]. Supported by recent investigations [6] we believe that amorphous materials, although metastable, are excellent candidates for corrosion resistant coatings.

Unless stated otherwise, etch rates in the following were determined by measuring film thickness using variable angle scanning ellipsometry (VASE) before and after exposure to the alkaline solutions. To be able to measure etch rates as low as 0.1 Å/h ($\approx 0.1$ μm/year) exposure times of several months are required.

The best diffusion barrier known from literature is probably amorphous silicon tantalum nitride (a-$Si_xTa_yN_{1-x-y}$) with $x = 0.14$ and $y = 0.36$ [7], combining the inertness of interstitial compounds with an effective blocking of grain boundary diffusion through its amorphicity. This alloy has also been demonstrated as a micromechanical construction material [8]. We investigated reactively co-sputtered a-$Si_xTa_yN_{1-x-y}$ using a Si and a Ta source in an $Ar/N_2$ ambient. The etch rate of a-$Si_xTa_yN_{1-x-y}$ in a KOH solution (pH 13) at 110 °C was strongly composition dependent and best results of 2 Å/h were obtained for a composition of $x \approx 0.17$ and $y \approx 0.33$, which is close to the composition of the above-mentioned diffusion barrier.

We recently reported [9] on our investigations of amorphous tantalum oxide (a-$Ta_xO_{1-x}$) with a composition close to that of stoichiometric tantalum pentoxide ($Ta_2O_5$). The a-$Ta_xO_{1-x}$ films ($x \approx 0.29$) were deposited by DC-reactive sputtering from a Ta target in an $Ar/O_2$ plasma. Films with a thickness of 500 nm exhibited fewer than 3 pinholes per $cm^2$. The a-$Ta_xO_{1-x}$ crystallized into $Ta_2O_5$ upon annealing in $O_2$ at temperatures higher than 550 °C. Whereas the crystallized films were released from the silicon substrate after only a few days of exposure to a pH 13 solution at 140 °C during accelerated testing, no change in thickness of the a-$Ta_xO_{1-x}$ was observed after exposure to the same solution for 3 months, yielding etch rates lower than 0.02 Å/h (see Fig. 5).

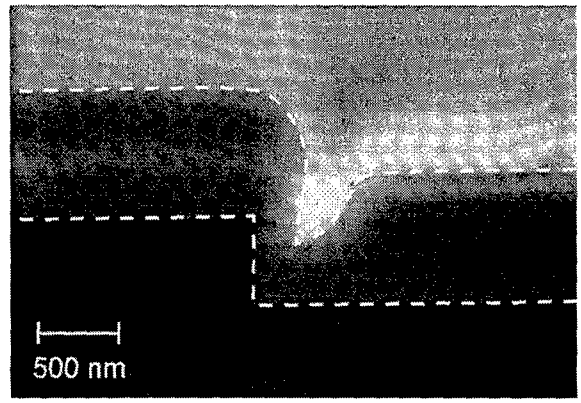



Fig. 6. Cross-sectional scanning electron micrograph showing step coverage of reactively sputtered a-$Ta_xO_{1-x}$ ($T_{dep} = 200$ °C) on an etched Si substrate. The dashed white lines indicate the profile of the a-$Ta_xO_{1-x}$ film.

Although the coverage of sharp corners and deep cavities anisotropically etched in Si(100) is good, the coating is not conformal, as typical for a sputtering deposition process. Such a non-conformal profile is shown in Fig. 6. For this particular case the angle between the substrate surface normal and the Ta target was 45°. If better conformity is required, a metal-organic chemical vapor deposition process may be the solution.

Another important feature of the a-$Ta_xO_{1-x}$ films is that the residual stress can be controlled by annealing. As can be seen in Fig. 7, the compressive stress of 200 MPa of samples deposited at RT can be eliminated by annealing 30 min in oxygen at temperatures slightly higher than 400 °C. This temperature is below the crystallization temperature of the amorphous oxide, thus maintaining the excellent etching and adhesion properties of the film. Other advantageous properties of a-$Ta_xO_{1-x}$ include the possibility of patterning the film by etching in concentrated hydrofluoric acid solutions, a high dielectric strength in excess of $3 \times 10^6$ V/cm, as well as biocompatibility.

Hydrogenated amorphous silicon carbide (a-SiC:H), obtained from third parties through a commercially available plasma-enhanced chemical vapour deposition (PECVD) process [10] appears to be an excellent coating as well. Initial

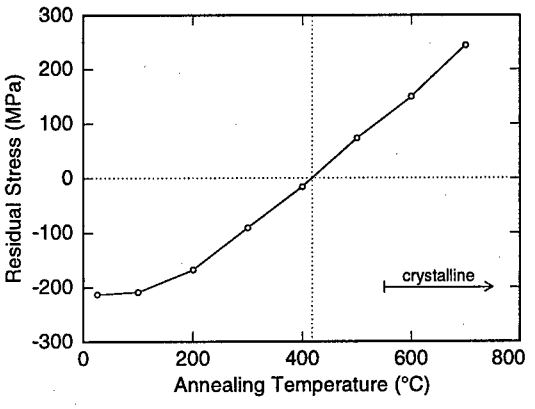



Fig. 7. Residual thin-film stress for a-$Ta_xO_{1-x}$ films ($T_{dep}$ = 20 °C) on Si(100) after post-annealing in 1 bar $O_2$ for 30 min at the temperatures indicated. After annealing at temperatures slightly above 400 °C stress-free films are obtained, whereas crystallization of a-$Ta_xO_{1-x}$ into $Ta_2O_5$ first occurs above 550 °C.



Fig. 8. Cross-sectional scanning electron micrograph showing perfect step coverage of a-SiC:H over a poly-Si interconnect on a Si substrate. The a-SiC:H is 0.2 μm thick and was deposited by PECVD.

measurements [4] showed that the etch rates in pH 11 at 75 °C were sufficiently low (see also Fig. 5). By modifying the PECVD process we were able to reduce the etch rate by more than two orders of magnitude. An etch rate of 0.01 Å/h at 140 °C was determined. The modified material is indicated in Fig. 5 as a-SiC:H(II).

The excellent conformal coating properties of the a-SiC:H PECVD process are depicted in Fig. 8, which shows a cross section of a-SiC:H over a poly-Si interconnect line. Details on this work will be published elsewhere [11].

A (large) number of diamond-like carbon (DLC) films was obtained from several vendors. Although the best DLC films did not etch detectably in pH 11 solutions at 140 °C, we observed, without exception, that the alkaline solution penetrates the film locally. The result is that the DLC

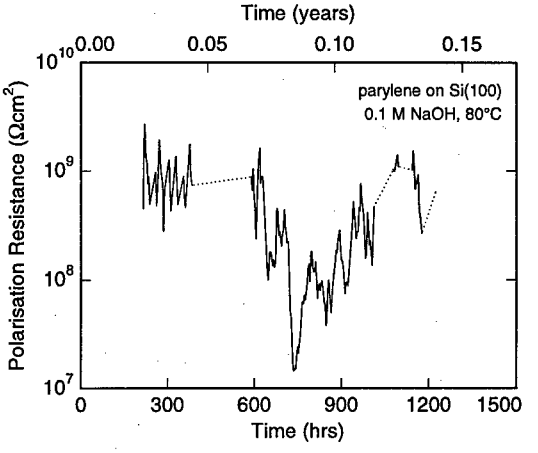



Fig. 9. Polarisation resistance of parylene C on Si(100) as a function of exposure time to a 0.1 M NaOH solution at 80 °C. A polarisation resistance higher than $10^9$ Ωcm$^2$ usually indicates 'no etching'.

films peel off from the silicon or oxidized silicon substrates within a couple of days.

Parylene C is well-known for its passivation behavior. An *in situ* determination of the etch rate of dielectric films was performed by electrochemical measurements similar to those presented by the Motorola group [12]. The ion flux through the coating relates directly to the polarisation resistance. A polarisation resistance higher than $10^9$Ωcm$^2$ represents totally inert films.

A typical polarisation resistance measurement of a parylene C coating in a 0.1 M NaOH solution at 80 °C is shown in Fig. 9. Although variations in the measurements occur, the resistance recovers to its initial value of about $10^9$ Ωcm, confirming the inertness of the parylene C films.

We observed no degradation of 4 μm thick parylene C films on Si after exposure to a 0.1 M NaOH solution at RT for three months. Similar films exposed to the same solution at 80 °C seem not to etch either, but after maximum three months the coating delaminates, most likely due to penetration of moisture at the interface. Similar undesired lift-off behavior was observed for polyimides. No etching was observed and the typical exposure time to failure is less than one week.

### 3.2 *Glue: chip mounting of differential pressure sensors*

We determined the properties of a number of glues and tested the suitability of the most promis-

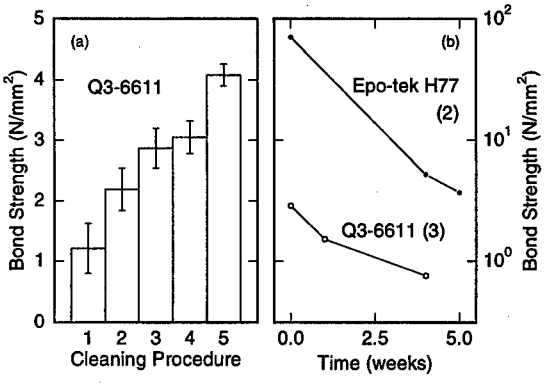



Fig. 10. (a) Bond strength (average of 10 measurements) of Q3-6611 glue directly after gluing for various surface cleaning procedures (see text). Note that the variation in bond strength becomes smaller with increasing strength. (b) Bond strength for Q3-6611 (cleaning procedure 3) and Epo-tek H77 (procedure 2) after exposure to a pH 11 solution at 95 °C as a function of exposure time.

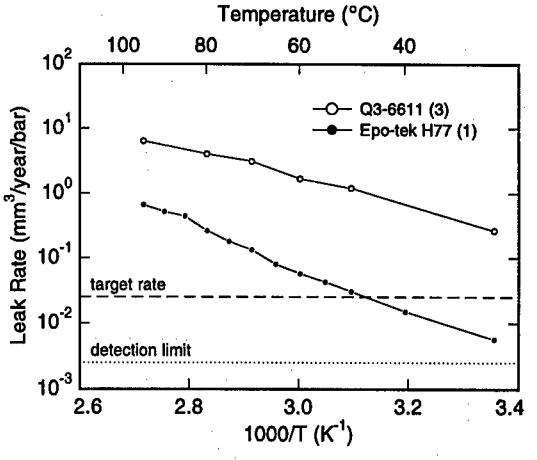



Fig. 11. Leak rates of Q3-6611 silicone glue and Epo-tek H77 epoxy. The Epo-tek H77 was exposed to a pH 11 solution at RT for 18 months prior to the measurements. The dashed line indicates an acceptable leak rate and the dotted line is the detection limit. Cleaning procedures are indicated in brackets.

ing types for chip mounting with the first concept presented in section 2 in mind (Fig. 1). The following glues were selected: Epo-tek H77 epoxy glue (Epoxy Technology Inc.), DM1110HT thermoplastic glue (DieMat), Semicosil 989 silicone glue (Wacker), and Q3-6611 silicone glue (Dow Corning).

Burst pressure tests were performed to determine the bond strength of various glues. $3 \times 4\,mm^2$ dies were glued onto a metal substrate over a hole with a diameter of 2 mm. Pressure was then applied through the hole until the dies were released from the substrate. The influence of the surface cleaning procedure on the bond strength of the Q3-6611 glue is shown in Fig. 10(a). The following cleaning procedures correspond to the numbers in Fig. 10(a):

(1) 2-propanol
(2) soap and 2-propanol
(3) primer
(4) soap, 2-propanol, and primer
(5) soap, 2-propanol, primer, and bake at 150 °C.

From Fig. 10(a) it can be seen that the standard primer for Q3-6611 (treatment 3) is highly effective and the bond strength is only slightly enhanced by soap and propanol cleaning prior to the priming treatment (4). A soap treatment prior to 2-propanol cleaning (2) strongly enhances the effect of 2-propanol cleaning only (1). However, additional baking (5) for several hours at the recommended curing temperature (150 °C) of the Q3-6611 glue yields the highest bond strengths ($> 4\ N/mm^2$). It is also noted that the variation in the bond strength is smaller for higher bond strengths.

Fig. 10(b) shows the degradation of the bond strength for Q3-6611 and Epo-tek H77 due to exposure to a pH 11 solution at 95 °C. The bond strength of the Q3-6611 silicone glue (cleaning procedure 3) is strongly reduced after a few weeks of exposure. Also the bond strength of the Epo-tek H77 (cleaning procedure 2) is reduced. After five weeks of exposure the bond strength has fallen from more than $70\ N/mm^2$ to less than $4\ N/mm^2$. The desired bond strength is of course dependent on the bond area and the application, but a minimum value of $4\ N/mm^2$ is usually required.

Moisture which penetrates the adhesive bond will eventually deteriorate the electrical circuitry in the sensor housing by corrosion. To ensure a sensor system lifetime of 10 years, a maximum leak rate of $2.5 \times 10^{-2}\ mm^3/year/bar$ is allowed. A highly sensitive test-rig was built in which the leak rate of heavy water is measured by quadrapole mass spectrometry with a detection limit of $2.5 \times 10^{-3}\ mm^3/year/bar$.

Chips ($5 \times 7.5\,mm^2$) were glued on a substrate with a hole (2 mm in diameter). The leak rate of Q3-6611, measured directly after gluing using preparation method 3, is higher than the acceptance level for all temperatures ranging from RT to

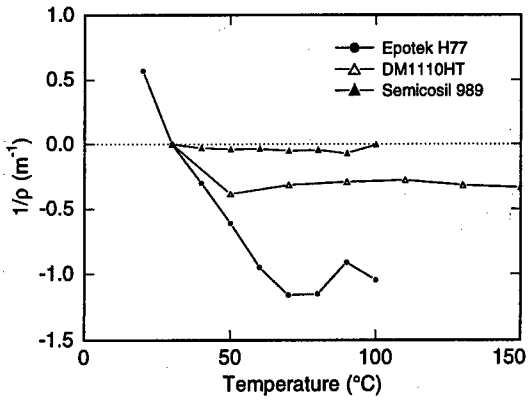



Fig. 12. Curvature $1/\rho$ of mounted Si chips on $Al_2O_3$ thick film substrates. The measurement at 30 °C is taken as reference. Glue types are indicated.

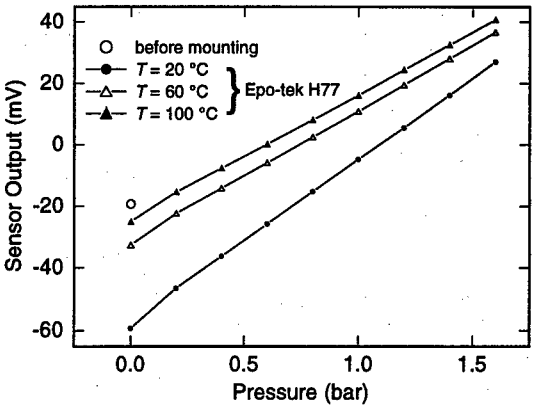



Fig. 13. Output of a piezoresistive pressure sensor before and after mounting using Epo-tek H77 (5 V bias applied). Gluing of the sensor introduces a large offset. We observe an offset drift with temperature change. The drift is strongly reduced for $T > 70$ °C, when the glue softens.

95 °C (see Fig. 11). The leak rate of water through Epo-tek H77 (cleaning procedure 1) is lower than the detection limit of our system. However, after exposure to a pH 11 solution at RT for 18 months, the leak rate of the Epo-tek H77 is above the detection limit but below the target rate for temperatures up to 50 °C (see Fig. 11). Improvements are required to fulfill our specifications.

Chip curvature measurements were performed using an optical surface profiler. Changes in the curvature are related to stress induced by the die attach. A typical situation, where a 350 $\mu$m thick silicon chip was glued onto a thick-film $Al_2O_3$ substrate, was investigated. The curvature, $1/\rho$, was determined from curvature difference measurements with the data collected at 30 °C as reference. In Fig. 12 the results are shown for three types of glue. A negative sign of $1/\rho$ corresponds to compressive stress. The soft silicone glue Semicosil 989 does not introduce considerable bending when the temperature is raised. The DM1110HT thermoplast is slightly harder and introduces additional chip bending for temperatures up to 50 °C, where the glue softens. The hard epoxy, Epo-tek H77, influences the chip curvature most with a change in $1/\rho$ of $-1.75$ m$^{-1}$ between RT and 70 °C. Above this temperature $1/\rho$ does not change significantly due to softening of the glue.

Not shown is the effect of humidity and moisture absorption in Epo-tek H77. Thin-film stress was measured by wafer curvature difference measurements of a 100 $\mu$m thick Epo-tek H77 layer on a silicon wafer. Absolute values of 8 MPa and 2 MPa (tensile stress) were measured for dry (annealed) samples and samples exposed to ambient for one week, respectively. The relaxation time of the moisture absorption process is a few days and the measurements are reversible in the temperature range RT–150 °C.

The influence of chip mounting using Epo-tek H77 glue on pressure sensor performance is shown in Fig. 13. The sensor system was as shown in Fig. 1. Before mounting a sensor offset of $-20$ mV (using a 5 V bias) was measured at RT (open circle). After mounting the offset changes to $-60$ mV (closed circles). An offset drift with temperature change is observed. The temperature drift is strongly reduced for temperatures above 70 °C. Below this temperature a small change in sensitivity occurs as well, as can be seen by the change in slope of the data collected at 20 °C and 60 °C, respectively. These observations are in agreement with the curvature measurements above, where we observed softening of the glue around 70 °C.

### 3.3 *Thin-film anodic wafer bonding*

Wafer bonding processes reduce the need for chip-scale handling and are used to actually form devices (such as the sensor shown in Fig. 2) as well as for back-end processing to encapsulate devices. Anodic bonding is an attractive method due to its relatively low required temperatures (< 450 °C) and silicon-to-glass bonding is frequently

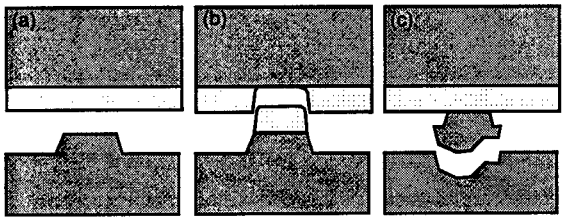



Fig. 14. Typical fracture behaviour of thin-film anodically bonded wafers. A structured wafer is bonded to a glass coated wafer (a) and then subjected to pull test. Fracture occurs either in the glass (b) or the silicon structure (c). Fracture at the bond interface is not observed.

applied (see e.g. Refs. [13] and [14]). Compared with silicon-to-glass bonding major advantages of silicon-to-silicon wafer bonding using thin films are (1) the elimination of differences in thermal expansion between the two wafers and (2) strongly enhanced functionality and the formation of truly 3-dimensional structures by joining highly micromachined and processed wafers.

The glass layers required for thin-film wafer bonding are usually deposited by sputtering [15]. We developed a bonding process for silicon to silicon using evaporated glass layers [16,17]. The deposition rates of the evaporated glass are approximately 20 nm/s (1.2 $\mu$m/min), which is three orders of magnitude higher than for sputter-deposited glass. This is in favor of industrial applications, since glass thicknesses of several microns are required for a reliable bonding process.

The typical fracture behaviour of anodically bonded structures using the evaporated glass is shown in Fig. 14. A cross section of a structured and a glass-coated wafer before bonding is shown in Fig. 14(a). After bonding and pull testing it is observed that fracture either occurs in the glass (Fig. 14b) or in the micromachined silicon structure (Fig. 14c). Fracture at the bond interface occurs rarely.

The bond strengths of the thin-film anodic bonding were determined by pull testing after bonding at various temperatures (see Fig. 15). Each data point is averaged over more than 50 samples from several bond experiments, resulting in a typical variation of 5 N/mm$^2$. We were not able to achieve bonding below 200 °C. From 200 °C to 300 °C the bond strength linearly increases to a value of approximately 25 N/mm$^2$ and saturates at higher bonding temperatures. Above 225 °C the bond yield is better than 90%.

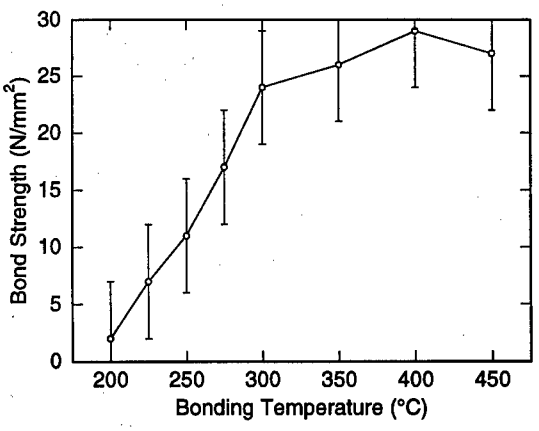



Fig. 15. Bond strength of the thin-film anodic bond as a function of bonding temperature. Bonding is achieved above 200 °C and the bond strength saturates around 300 °C. Each point is averaged over more than 50 samples from several bonding experiments. The typical standard deviation is 5 N/mm$^2$.

Further, we have shown [18] that not only residual stress in the glass layer, but also small temperature gradients during bonding on the order of a few degrees centigrade strongly influence the curvature of the bonded structures.

### *3.4 Feedthroughs*

To obtain the hermetic seals required for mounting of the needle sensor (see Fig. 3) we used glass frits which were customized to match the thermal expansion of silicon. Temperatures of 650 °C were required to obtain hermetic seals. This temperature is much higher than the eutectic temperature of the glass frit and is incompatible with most metallization schemes. The mounted sensors withstood burst pressure tests up to 4000 bar. Furthermore, we subjected the mounted sensors to thermal shock test. After 30 000 cycles from −40 to 120 °C no leakage was observed at RT.

A novel interconnection technique compatible with standard semiconductor processes was developed [19,20]. The technique uses electrodeposited photoresist (Eagle 2100ED by Shipley Europe Ltd.) and provides high vertical wiring densities for wafer through-hole interconnects and into deep cavities. An example of the through-hole interconnects is shown in Fig. 16. The size of the holes in the 350 $\mu$m thick Si(100) wafer is 1 × 1 mm$^2$. The wires consist of Ti/Au with

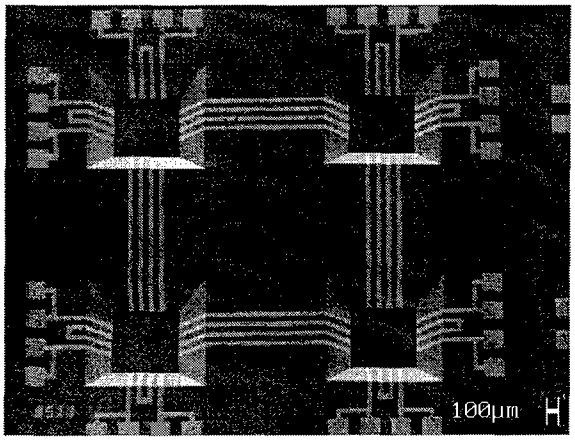



Fig. 16. Multiple vertical electrical feedthroughs. The size of the holes is $1 \times 1$ mm$^2$. The line width of the Ti/Au (20 nm/400 nm) wires is 50 $\mu$m.

thicknesses of 20 nm/400 nm and are 50$\mu$m wide. Reliable through-hole electrical feedthroughs were obtained for line widths down to 20 $\mu$m and a density of 250 wires per cm.

## 4 Conclusions and perspectives

Three concepts for micromachined silicon-based sensors exposed to harsh media are presented and their required technologies were investigated.

Innovative materials are required since conventional chemical resistant layers such as $Si_3N_4$ do not offer sufficient protection against highly alkaline environments. The protective coatings *a*-$Ta_xO_{1-x}$ and *a*-SiC:H fulfill the requirements for corrosion resistance of 1 $\mu$m per 10 years in a pH 11 solution at 125 °C with etch rates around $10^{-2}$ Å/h. *a*-$Si_xTa_yN_{1-x-y}$, an excellent amorphous diffusion barrier in metallization schemes, exhibits etch rates of approximately 2 Å/h in a narrow composition range around $x = 0.17$ and $y = 0.33$. DLC films apparently do not etch, but delaminate due to local penetration (through pinholes) of the etchant. Delamination after exposure to the media for several hours to days, is the typical failure mechanism we observed for crystalline $Ta_2O_5$ and a number of polymer coatings, including parylene C.

Several types of glue and their suitability as die attach were investigated. The bond strength of Q3-6611 silicone glue can be improved by extended curing procedures. Bond strengths of Q3-6611 and Epo-tek H77 epoxy are strongly reduced after exposure to pH 11 solutions at 95 °C for several weeks. Leak rates of Q3-6611 are on the order of 1 mm$^3$/year/bar. Leak rates of Epo-tek H77 are below the detection limit of $2.5 \times 10^{-3}$ mm$^3$/year/bar directly after gluing and increased after exposure to a pH 11 solution at RT for 18 months. The target leak rate of $2.5 \times 10^{-2}$ mm$^3$/year/bar was obtained below 50 °C. The Epo-tek H77 softens around 70 °C, which is also the temperature above which sensor offset drift is strongly reduced. Glass frits were successfully applied to hermetic sealing of the needle sensor.

Furthermore, we developed a thin-film anodic wafer bonding process using evaporated glass with high deposition rates (20 nm/s) and bond strength ($> 25$ N/mm$^2$) for bonding temperatures above 300 °C. Through-hole electrical feedthroughs were obtained using electro-depositable photoresist. Wire densities of 250 wires per cm and line widths down to 20 $\mu$m were obtained.

It is expected that a combination of technologies presented in this paper leads to generic total packaging solutions for sensors exposed directly to harsh environments.

## Acknowledgements

The authors greatly acknowledge Prof. Dr. K. Najafi (Mikroelektronik Centret, on leave from the University of Michigan, Ann Arbor, U.S.A.) for his critical reading of the manuscript. P.E. Andersen (Grundfos A/S) is thanked for the many helpful discussions. Technical support was provided by P.N. Egginton and F. Jensen (Mikroelektronik Centret), and P. Brandt (Danfoss A/S). M. Lindahl, J. Kuhmann (Mikroelektronik Centret), P. Kersten (present address: Photo Print Electronic GmbH, Schopfheim, Germany) and S. Henke (present address: Angewandte Display-Technologie GmbH, Stuttgart, Germany) participated at an early stage of the project. We would like to thank Dr. S. Gasser and Prof. Dr. M.-A. Nicolet (Caltech, Pasadena, U.S.A.) and Dr. W. Olthuis (MESA Institute, Technical University of Twente, The Netherlands) for providing us with the first Si–Ta–N and Ta–O test samples, respectively. This work was supported by the 'Materials for Advanced Micromechanical Packaging' program, under the Materials Development Program supported by the Danish Agency for

Trade and Industry, The Danish Natural Science Research Foundation, and the Danish Technical Science Research Foundation.

## References

[1] Dyrbye K, Brown TR, and Friis Eriksen G. Packaging of physical sensors for aggresive media applications. J. Micromech. Microeng. 6 (1996) 187–192. Invited paper at Micromechanics Europe 1995 (September 1995, Copenhagen, Denmark).

[2] Jensen OS and Gravesen P. A new piezoresistive pressure transducer principle with improvements in media compatibility. J. Micromech. Microeng. 6 (1996) 105–107.

[3] Friis Eriksen G, Dyrbye K, Weichel S, and de Reus R. An SOI-based flat surface pressure sensor. In preparation.

[4] Eriksen GF and Dyrbye K. Protective coatings in harsh environments. J. Micromech. Microeng. 6 (1996) 55–57.

[5] de Reus R. Thin-film diffusion barriers. In: Westbrook JH and Fleischer RL (eds), *Intermetallic Compounds—Principles and Practice*, vol 2: Practice, chap 29, pp 603–635. John Wiley & Sons Ltd., Chicester, UK, 1994.

[6] Andersen P. Protection of micromechanical pressure sensors with metallic glasses. Ph.d. thesis, University of Aarhus, Århus, Denmark, 1998.

[7] Kolawa E, Chen JS, Reid JS, Pokela PJ, and Nicolet MA. Tantalum-based diffusion barriers in Si/Cu VLSI metallizations. J. Appl. Phys. 70 (1991) 1369–1373.

[8] Linder C, Dommann A, Staufert G, and Nicolet MA. Amorphous metals as new materials for transducer applications. In: Puers R (ed), *Proceedings of Eurosensors X, the 10$^{th}$ European Conference on Solid-State Transducers*, pp 473–476. Leuven, Begium, September 1996.

[9] Christensen C, de Reus R, and Bouwstra S. Tantalum oxide thin films as protective coatings for sensors. In: *MME'98 Workshop Digest*, pp 248–251. SINTEF, Ulvik in Hardanger, Norway, June 1998. Also submitted to J. Micromech. Microeng.

[10] Klumpp A, Schaber U, Offereins H, Kühl K, and Sandmeier H. Amorphous silicon carbide and its application in silicon micromachining. Sensors and Act. A 41–42 (1994) 310–316.

[11] Friis Eriksen G, Dyrbye K, and Romedahl Brown T. Silicon carbide protective coatings for sensors in alkaline solutions. In preparation.

[12] Bitko G, Monk D, Maudie T, Stanerson D, Wertz J, Matkin J, and Petrovic S. Analytical techniques for examining reliability and failure mechanisms of barrier coating encapsulated silicon pressure sensors exposed to harsh media. In: Chau K and Roop R (eds), *Micromachined devices and components II*, vol 2882 of *Proc. SPIE*, pp 248–258, 1996.

[13] Esashi M. Encapsulated micromechanical sensors. Microsystem Technol. 1 (1994) 2–9.

[14] Dokmeci M, Von Arx J, and Najafi K. Accelerated testing of anodically bonded glass-silicon packages in salt water. In: *1997 International Conference on Solid-State Sensors and Actuators (Transducers '97)*, vol 1, pp 283–286. IEEE Electron Devices Society, June 1997. Chicago, Illinois, U.S.A.

[15] Hanneborg A, Nese M, and Øhlckers P. Silicon-to-silicon anodic bonding with a borosilicate glass layer. J. Micromech. Microeng. 1 (1991) 139–144.

[16] de Reus R and Lindahl M. Si-to-Si wafer bonding using evaporated glass. In: *1997 International Conference on Solid-State Sensors and Actuators (Transducers '97)*, vol 1, pp 661–664. IEEE Electron Devices Society, June 1997. Chicago, Illinois, U.S.A.

[17] Weichel S, de Reus R, and Lindahl M. Silicon-to-silicon wafer bonding using evaporated glass. Sensors and Act. A (1998). In press.

[18] Weichel S, de Reus R, and Bouwstra S. Residual stress in thin-film anodic bonding. In: *MME'98 Workshop Digest*, pp 47–50. SINTEF, Ulvik in Hardanger, Norway, June 1998. Also submitted to J. Micromech. Microeng.

[19] Christensen C, Kersten P, Henke S, and Bouwstra S. Wafer through-hole interconnections with high vertical wiring densities. IEEE Trans. Components Packag. Manufacturing Technol. A 19 (1996) 516–522.

[20] Kersten P, Bouwstra S, and Petersen J. Photolithography on micromachined 3D surfaces using electro-deposited photoresist. Sensors and Act. A 51 (1995) 51–54.

PERGAMON

Microelectronics Reliability 38 (1998) 1261–1264

MICROELECTRONICS RELIABILITY

# Analysis of thermomechanical stresses in a 3D packaged micro electro mechanical system

C. Pellet [a], M. Lecouvé [a], H. Frémont [a], A. Val [b], D. Estève [b]

[a] *IXL, UMR CNRS 5818, Université Bordeaux I, 351 cours de la libération, 33405 Talence Cedex ,France*
[b] *LAAS , 7 avenue du colonel Roche, 31077 Toulouse Cedex*

**Abstract**

Temperature cycling has been used associated to FEM simulations to evaluate thermomechanical stresses in a MCM-V packaged MEMS. The places of maximum stresses have been localized by the simulations. The results are coherent with electrical measurements done on the assembly during and after the thermal cycling. Modifications are proposed (type of material , geometry of the package) in order to improve the reliability. © 1998 Elsevier Science Ltd. All rights reserved.

## 1 . Introduction

Micro Electro Mechanical System is an emerging technology with demonstrated potential for a wide range of applications including sensors and/or actuators and their electronic circuits for medical, industrial, consumer, automotive and instrumentation products. MEMS includes weak structures as diaphragms, cantilevers,... Special care has to be taken for their assembly. Since monolithic integration of MEMS is not always possible, MCM assembly is often used. The MEMS studied here was packaged using the MCM-V assembling technique developed by 3Dplus-THOMSON/CSF companies [1]. A section of the 3D assembly is represented on the figure 1. This assembly, developed for an European project (BARMINT) [2], includes a micropump and silicon chips. The micropump is realized by a silicon diaphragm and an heater unit composed by a resistor deposited on an alumina substrate. Since different materials are present in the MEMS the main problem comes from the mismatch between the molding compound and the other materials coefficient of thermal expansion (CTE). This mismatch induces thermomechanical stresses and strain within the packaged MEMS particularly at the interfaces during temperature cycles and the post molding cooldown of the assembly. The induced stresses and strain can damage the silicon diaphragm as well as the electronic chips. This paper reports the evaluation of thermomechanical stresses inside the assembly by means of thermal cycling associated to finite element simulations. A specific assembly was used as test vehicle for these experiments [3]. It consists in an micropump and four silicon chips on which electrical lines have been deposited, according to the figure 2. Two families of electrical lines were used. The first one consists in three lines of different widths (10 μm, 25 μm, 45 μm) located at the edge of the chip. The second one consists in three lines of different widths (10 μm, 25 μm, 45 μm) located at the center of the chip. The different resistors deposited on each silicon chip are accessible through contact points along the 3D cube faces. Electrical tests have been implemented before, during and after the thermal cycling. Visual inspection and X-ray analyses have also been used.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00103-6

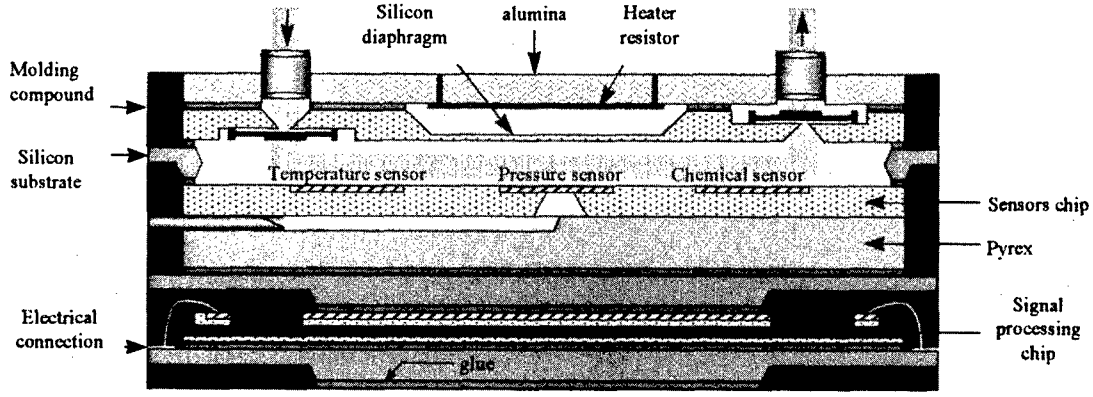



Fig. 1 : Schematic representation of the MEMS

## 2 . Temperature cycling

Temperature cycling was performed from -50°C to 150°C. The cycle period is 4 hours 36 minutes. The decreasing and increasing speeds are lower than 7°C/min to ensure that the whole assembly is always maintained at the same temperature. Ten devices were electrically tested and observed after 24 hours, one week and 3 weeks. Contact losses were already observed after 24 hours thermal cycling. Therefore the electrical tests have revealed differences between electrical lines and also between the silicon chips according to their location inside the package. Few contact losses occur for electrical lines located in the middle of a chip. On the contrary a high yield of contact losses was obtained for the lines located at the edge of a chip. Concerning the different chips, the chip just beneath the micropump (level 1) is the more affected while the two chips located in the middle of the assembly (level 2 and 3) are the less affected ones (see Fig. 3).

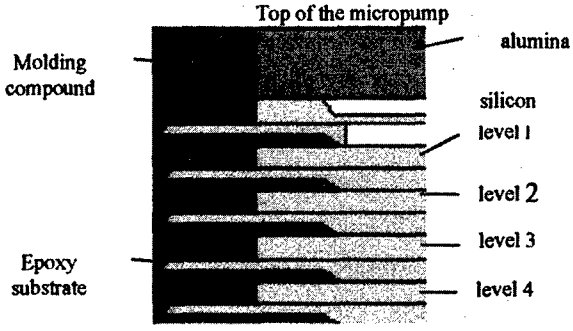



Fig. 2 : Section of the test vehicle

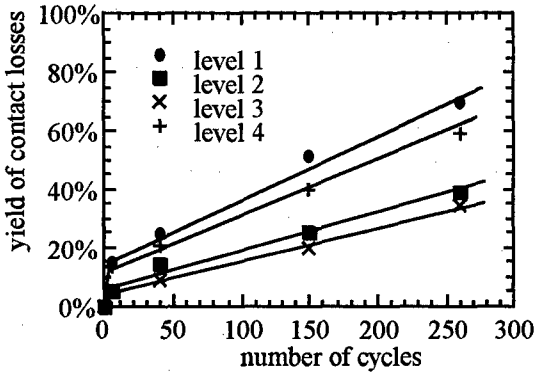



Fig. 3 : Yield of contact losses as a function of cycle number

## 3 . Finite element simulations

Two dimensional finite element model of the MEMS was constructed and analyzed on a SUN station using ANSYS 5.1 software. In this model only four materials are taking into account : molding compound, silicon, alumina and epoxy substrate, the adhesives are neglected as well as the metal lines and connections. The package is supposed undeformed and unstressed at room temperature. The analysis is carried out to simulate the geographical distribution of the stess along the principal axis as a function of temperature. During the decreasing or the increasing of the temperature one consider that all materials have a elastic behavior except for the molding compound. For this material the Young modulus depends of the

temperature varying from 1.5 $10^4$ MPa at –50°C to 0.8 $10^4$ MPa at 140°C. The other materials properties are reported in table 1.

Table 1
Materials properties

| Material | Young modulus (MPa) | Poisson factor | Coefficient of Thermal Expansion |
|---|---|---|---|
| Silicon | 1.69 $10^5$ | 0.33 | 2.4 $10^{-6}$ |
| Alumina | 2.6 $10^5$ | 0.23 | 6.5 $10^{-6}$ |
| Epoxy | 4.3 $10^3$ | 0.278 | 15 $10^{-6}$ |
| Molding compound | - | 0.33 | 15 $10^{-6}$ |

Figure 4 presents the stress $\sigma_{xx}$ within the assembly at temperature of 140°C. Since the package is considered as symetrical only the half part of the system is represented. In the first column of the table 2 are reported the stresses $\sigma_{xx}$ in the different parts of the assembly at -50°C and 140°C. The stresses observed in the molding compound are higher than in the epoxy substrate while these materials have the same CTE because of the difference between their Young modulus. The stresses in the silicon chips is very high because of the low CTE of this material. They are especially high at the edge of the silicon chips. The silicon diaphragm experiences the maximum stress (-43 MPa (compressive stress) at -50°C, 65 MPa (extensive stress) at 140°C).

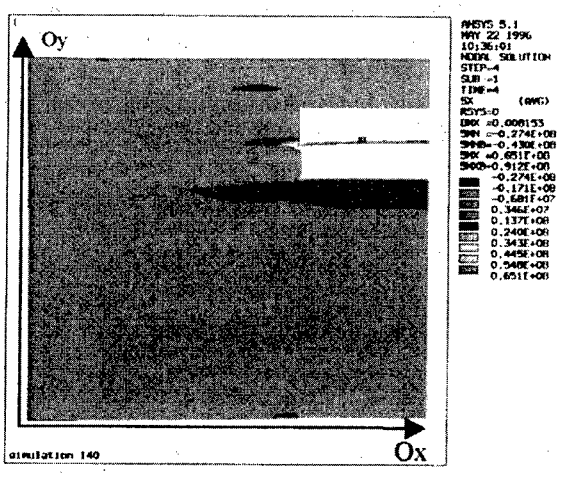



Fig. 4 : distribution of the stress intensity $\sigma_{xx}$ in the whole assembly

Table 2
$\sigma_{xx}$ stress (MPa) in the different parts of the assembly for the different material combination

| | | Real assembly | Assembly 2 | Assembly 3 |
|---|---|---|---|---|
| Top of the micropump | -50°C | 10 to 18 | 3 to 7 | 10 to 18 |
| | +140°C | -7 to -17 | -1.5 to -2.5 | -13 to -23 |
| silicon diaphragm | -50°C | -43 | -5 to -9 | -21 to -29 |
| | +140°C | 65 | 5 to 2 | 47 to 57 |
| silicon chip | -50°C | -20 | -5 to -13 | -13 to -37 |
| | +140°C | 13 to 27 | 5 to 14 | 7 to 27 |

## 4. Discussion

The thermomechanical stresses derived from the simulations are coherent with the results obtained during temperature cycling. The silicon chip located near the micropump (level 1) experiences more stress and strain than the other chips. It is also the case for the chip located at the base of the assembly (level 4). For one chip the place of maximum stress is the edge of the chip. Again it is coherent with the electrical measurements. FEM simulations also show that the silicon diaphragm experiences the maximum stress. However, X ray analyses have not allowed to localize contact losses but they have shown that the silicon diaphragm wasn't broken.

The finite-element model is hence validated and can be used to investigate other material combinations or geometries with the aim to reduce stress in the assembly. Two structures have been studied. The first one (denoted assembly 2) in which the alumina is replaced by silicon and a second one (denoted assembly 3) in which alumina and the epoxy substrates have been replaced by silicon. In the table 2 are reported the stresses $\sigma_{xx}$ in the different parts of the assembly for the different material combination. When the alumina is replaced by silicon the stresses decreased by a factor 3 especially for the silicon diaphragm and the chip located just beneath the micropump. There is no improvement for the last assembly where all the parts consist in silicon. The best material combination appears so to be the second one. This is probably due to the plastically influence of the epoxy which acts as a buffer layer between the silicon chips and the molding compound.

## 5. Conclusion

MEMS reliability as well as electronic component reliability largely depend on the packaging reliability [4]. Analysis of thermomechanical stresses in a 3D packaged MEMS shows that the choice of material used is a key factor for an improved reliability. Validated by experimental results FEM simulations are a powerfull tool since they allow to evaluate quickly different geometry or materials combination [5]. The study presented here has permited to demonstrate that 3D plastic encapsulation of a MEMS including a silicon diaphragm is possible without damaging the suspended structure.

## References :

[1] : C. Val, P. Bonnefoy, I. Drouin, Onde Electrique, mars-avril 1994, vol. 74, n° 2.
[2] : Esprit program, 1994-1997, Basic Research for Microsystems Integration"BARMINT".
[3] : A. Val, Etude de la technologie d'interconnexion verticale MCM-V appliquée aux microsystèmes, Thesis, Université Paul Sabatier- Toulouse, 1996.
[4] : O. Slattery, T. Hayes, W. Lawton, G. Kelly, C. Lyden, J. Barrett, C. O'Mathuna, J. of Materials Processing Technology, 1995, 54, pp 199-204.
[5] : C. Ducos, E. Saint Christophe, H. Frémont, G. N'Kaoua, C. Pellet, Y. Danto, Proc. Of European Symposium on Reliability of Electronic Devices (Ed. ELSEVIER), ESREF'97, 1997.

PERGAMON

Microelectronics Reliability 38 (1998) 1265–1269

MICROELECTRONICS RELIABILITY

# Electrical characterization and modification of a MicroElectroMechanical System (MEMS) for extended mechanical reliability and fatigue testing

D. Meunier [a,b,c], R. Desplats [a,b], J. Benbrik [a,b], G. Perez [a]
C. Pellet [b], D. Etsève [c], B. Benteo [d]

[a] *CNES, Toulouse, France*
[b] *IXL Laboratory, Bordeaux, France*
[c] *LAAS, Toulouse, France*
[d] *SOREP, Toulouse, France*

**Abstract**

The feasibility of microsystems has been achieved in the early 90's and is now followed by a growing industrialization period. Nevertheless, diversity of fields combined and materials heterogeneity generate new failure mechanisms different from traditional microelectronics. Now, if we take into account the lessons from the past in microelectronics, we note that failure analysis played a major role not only in development time reduction but also in qualification and reliability evaluation. That is why we decided to elaborate a specific failure analysis methodology for microsystems using existing instrumentation from microelectronics in order to prepare for present and future expert appraisement requirements in the micro-technologies field.

For this, we chose to perform accelerated aging tests of a commercial micro-accelerometer. This paper demonstrates how we realized the reverse engineering of this MicroElectroMechanical System (MEMS) using an Electron Beam Tester (Integrated Diagnostic System : IDS-5000) and how we modified its functions with a Focused Ion Beam (FIB-P2X) in order to control its full mechanical capabilities with electrical stimulation.
© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

New reliability problems appear with the emergence of microsystems. Even if the technological processes are close to those of microelectronic integrated circuits, the presence of micro-mechanical structures generates failure mechanisms different from traditional microelectronics. Now, there is still a real lack of knowledge of these new mechanisms, so that we chose to perform accelerated aging tests of a commercial micro-accelerometer in order to identify specific MEMS failure modes. ADXL50 from Analog Devices [1,2] is a complete acceleration measurement system (± 50G) on a single monolithic integrated circuit now used for airbag deployment.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00096-1

## 2. Operation description of ADXL50

ADXL50 is a BiCMOS silicon surface micromachined device. The polysilicon structure is a differential capacitor sensor which consists of 42 unit cells composed of independent fixed cantilevers and a movable floating cantilever beam. The two capacitors are series connected forming a capacitive divider in which the movable center plate deflects in response to changes in relative motion (Fig. 1).

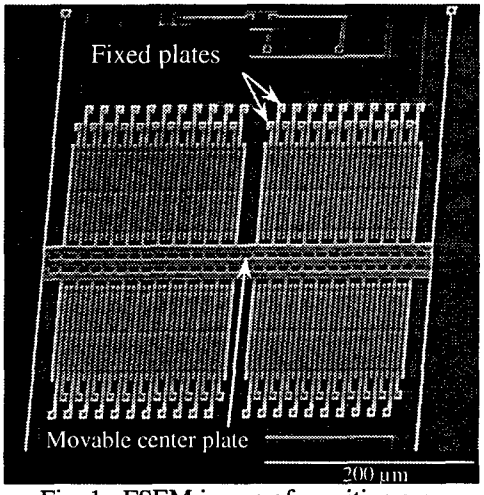



Fig. 1 : FSEM image of sensitive area

The two sensors' fixed capacitor plates are driven by 1 Mhz square waves of the same amplitude but 180° out of phase from one another. When at rest, the values of the two capacitors are the same and the output voltage at the center plate is zero (Fig. 2).

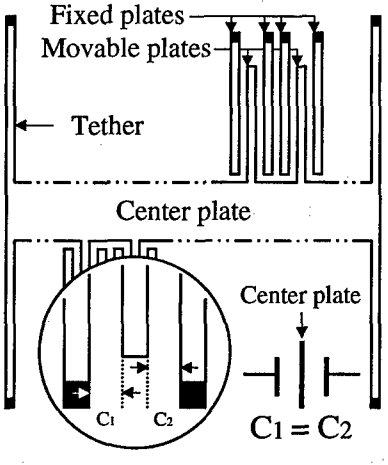



Fig. 2 : sensor at rest

When the sensor begins to move, a mismatch in the value of the capacitance is created producing an output signal at the center plate that will increase with the amount of acceleration (Fig. 3).

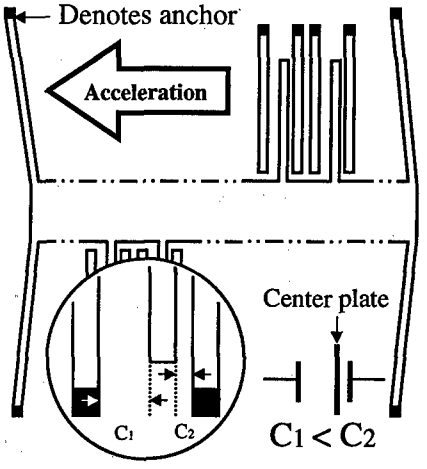



Fig. 3 : sensor responding to an externally applied acceleration

Information concerning the direction of beam motion is contained in the phase of the signal and is extracted with synchronous demodulation. Then the amplitude demodulator drives an amplifier providing a DC output voltage proportional to applied acceleration. This output stage is referenced to $V_{Ref}$ /2 (output voltage at 0G) and fed back to the sensor in order to force electrostatically the center plate to its original position when submitted to an acceleration (Fig. 4).

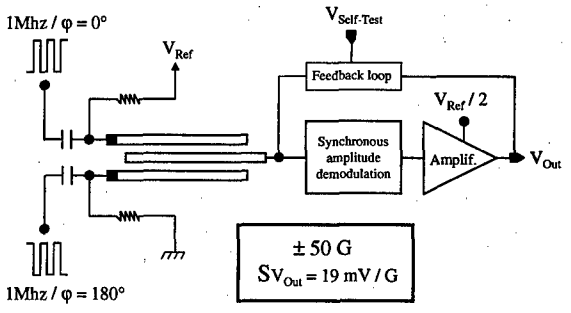



Fig. 4 : simplified electrical diagram

A self-test is initiated by applying a high-level voltage to the self-test input pin which causes the chip to apply an electrostatic force to the center plate. This function was designed to test the well operation and causes the sensor to deflect to the negative full-scale (-50G).

It is possible to activate mechanically the sensors' micro-parts by electrostatic attraction using an electrical stimulation. Nevertheless the self-test function only allows the maximum negative deflection position (-50G) and the self-test bandwidth is limited to a few kHz. That is why we decided to modify the circuit in order to simulate freely accelerations beyond the full nominal deflection scale.

## 3. Reverse engineering of the self-test function

First of all we proceeded to the reverse engineering of the self-test function using Electron Beam Tester IDS-5000 to identify inputs, outputs and alimentation lines ($V_{Ref}$ , $V_{Alim}$ and $V_{Ground}$). Signals were applied to the pins, so that we visualized the electrical states of the different lines using voltage contrast imaging and electron beam waveform measurement (Fig. 5,6,7).

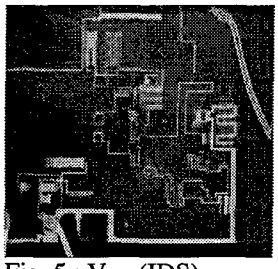


Fig. 5 : $V_{Ref}$ (IDS)

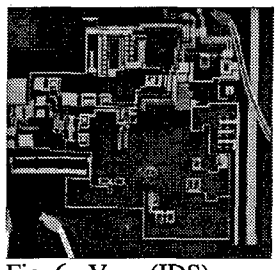


Fig. 6 : $V_{Alim}$ (IDS)

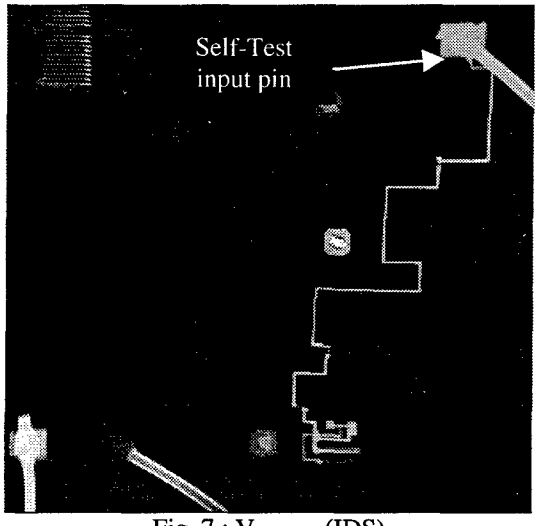



Fig. 7 : $V_{Self\text{-}Test}$ (IDS)

After what we completed this analysis with optical microscopy to extract the different intrinsic components (transistors, capacitors, resistors). Then we reconstitute a partial layout of the self-test area using semi-automatic image processing algorithm and we extracted the theoretical electrical diagram of the self-test function (Fig. 8).

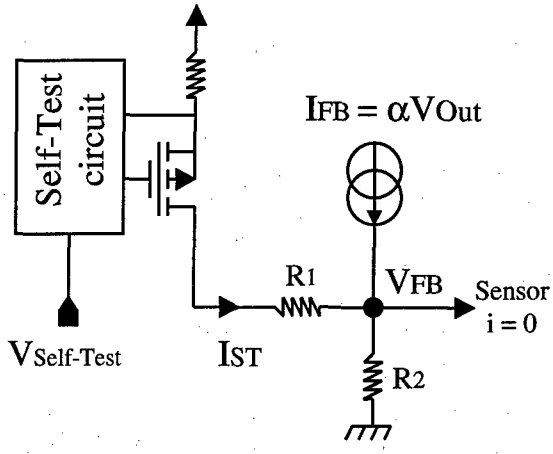



Fig. 8 : electrical diagram of the self-test function

The self-test circuit operates as a current source ($I_{ST}$) superposed to the feed-back current ($I_{FB}$) through $R_2$. Voltage increase at $V_{FB}$ involves electrostatic attraction of the center plate to its maximum negative position (-50G) providing a – 1 Volts change at $V_{Out}$ for a functional circuit.

## 4. Circuit modification using FIB-P2X

Our idea was to short-circuit the self-test function to connect its output to the self-test pin, so that we were able to apply easily any positive or negative external current using a HP4145 tester (Fig. 9).

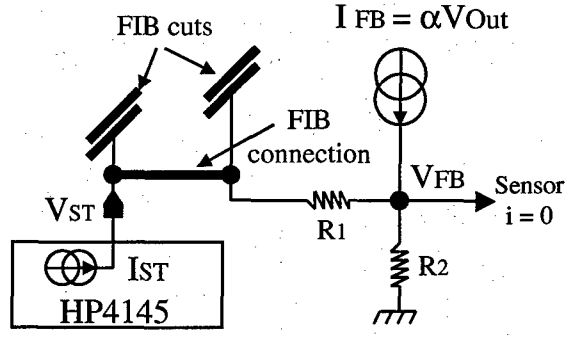



Fig. 9 : electrical diagram of FIB modification

We chose to modify the chip design close to the self-test function and far from the sensor area for two reasons :

1) To conserve main functionalities of the circuit between the original self-test current source and the sensor.
2) The polysilicon structure is very sensitive to the $Ga^+$ ion beam introduced by the FIB.

After we found the most appropriate chip design point we disconnected the self-test lines using dielectric and metal etching , then we short-circuited it using dielectric etching and metal deposition (Fig. 10). The circuit was powered during FIB operation to control in situ the well working of each step. In the end, no drifts appeared at the output voltage after FIB modification.

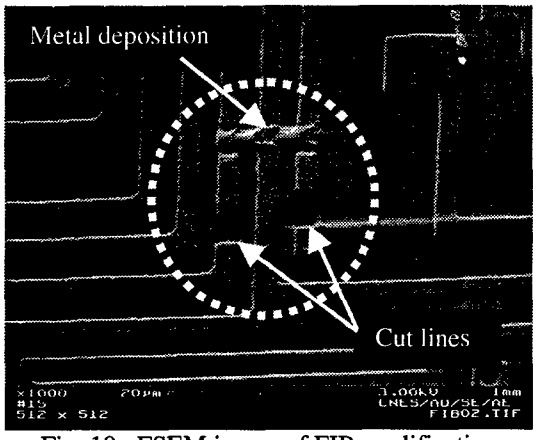



Fig. 10 : FSEM image of FIB modification

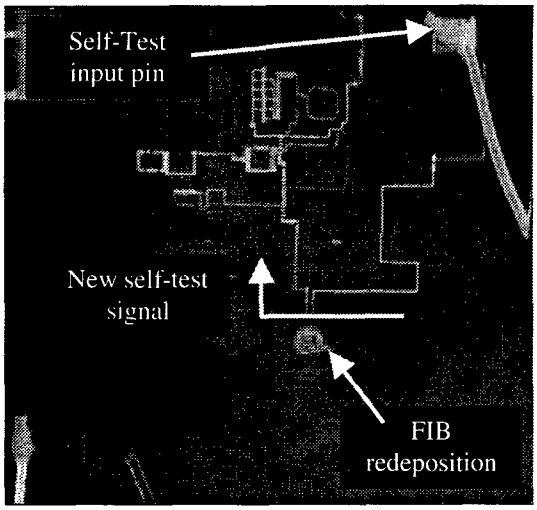


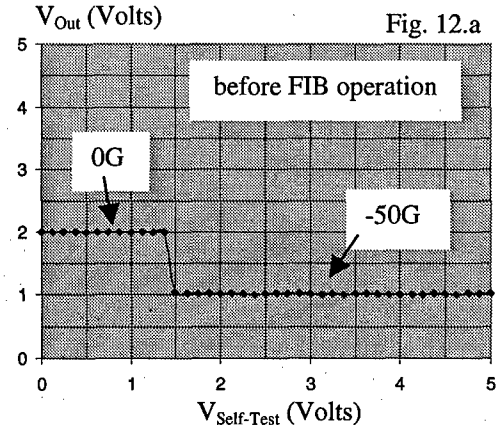

Fig. 11 : voltage contrast image of self-test signal

## 5. Results and discussion

### 5.1. Testing procedure

Measurements were realized on a HP4145 tester in ambient temperature and humidity conditions. ADXL50 is packaged in a hermetic 10-pin TO100 metal can. No drifts were observed during decapsulation stress or permittivity changes (Nitrogen → Air). Nevertheless each tests were performed with the metal can maintained on the chips' body because of its sensitivity to photons.

### 5.2. Measurements



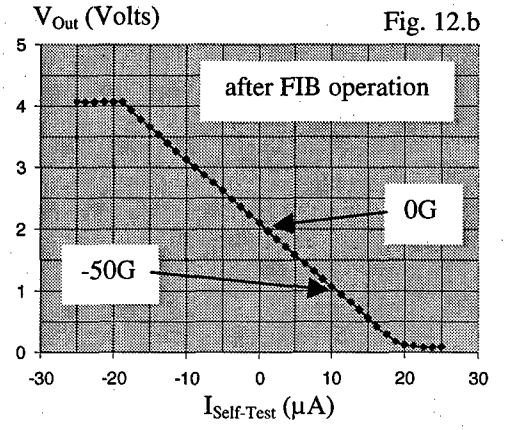




Fig. 12.a and 12.b : transfer characteristics

### 5.3. Discussion

Referring to figure 12.b, we succeeded not only in extending the acceleration range in positive values but also in increasing the deflection capabilities from approximately –100G to +100G. Indeed we are able to apply more than 100G accelerations due to the

control of the electrostatic voltage at the center plate, but these are not detectable because of the output stage voltage saturation.

First stress tests have been achieved involving problems of irreversible electrostatic sticking between the movable cantilever beam and the fixed cantilevers (Fig. 13). This failure mode is due to an excessive voltage applied to the center plate.

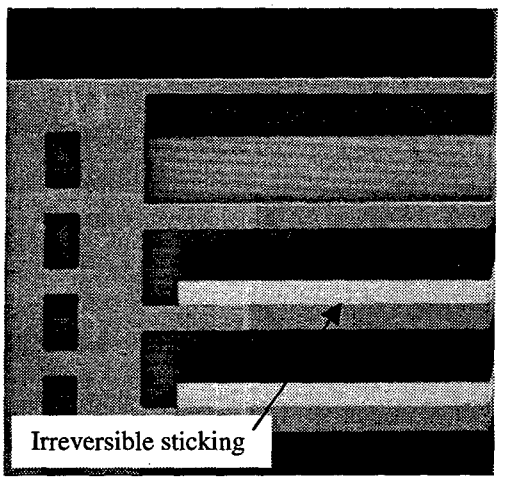



Fig. 13 : IDS voltage contrast image

## 6. Conclusion

Now are able to simulate shocks and vibratory environment. Further tests and mechanical simulation using ANSYS are currently undergoing in order to determine resonance frequency of the structure and to correlate electrical loading with mechanical loading.

## Acknowledgments

The authors acknowledge the contributions of Frederic Courtade, Bernard Baradat, Marylin Olle (technological analysis), and Michel Dupire (tests).

## References

[1] « Monolithic accelerometer with signal conditionning ADXL50 » Analog Device DataSheet

[2] S.J. Sherman, W.K. Tsang, T.A. Core, R.S. Payne, D.E. Quinn, K.H-L. Chau, J.A. Farash, S.K. Baum « A low cost monolithic accelerometer : product/technology update » IEEE proceedings 1992

PERGAMON

Microelectronics Reliability 38 (1998) 1271–1276

MICROELECTRONICS RELIABILITY

# Structures for piezoresistive measurement of package induced stress in transfer molded silicon pressure sensors

J. B. Nysæther[a,c], A. Larsen[b], B. Liverød[b] and P. Ohlckers[a,b]

*a Institute of Physics, University of Oslo, 0316 Oslo, Norway,*
*b SensoNor, P.O. Box 196, 3192 Horten, Norway*
*c SINTEF Electronics and Cybernetics, Postboks 124 Blindern, 0314 Oslo, Norway*

## Abstract

The package induced stress in a transfer molded micromechanical pressure sensor has been measured by use of silicon piezoresistors. Measurements have been carried out by monitoring the individual piezoresistors in the measuring bridge of an existing sensor, and by measurements on a modified sensor structure with specially designed stress sensing structures. The induced difference ($\sigma_x$ - $\sigma_y$) between x- and y- normal stress in the sensor diaphragm has been measured to 50 ± 10 MPa at 25°C. The induced stress creates a thermal zero shift of the sensor signal. By covering the sensor chip with a soft glob top before molding, the stress may be considerably reduced. For glob-topped samples the measured value of ($\sigma_x$ - $\sigma_y$) at 25°C is 20 ± 20 MPa.
© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Silicon piezoresistors are widely used as stress sensing elements in micromechanical devices such as pressure sensors and accelerometers. Piezoresistors can also be used for measuring so-called package stress [1,2]. Package stress is defined as the parasitic thermomechanical stress that arises because of the thermal expansion mismatch between the silicon based sensing element and the packaging materials. This temperature dependent stress may influence on the functionality or lifetime of the sensor element, and to improve the performance and reliability of the packaged device it is important to keep the package stress under control [1,3].

This is especially important for epoxy transfer molded devices, where the silicon based sensing element is encapsulated in a rigid polymeric material with a much higher thermal expansion. For transfer molded piezoresistive pressure sensors the large thermal mismatch may induce high levels of mechanical stress in the sensor diaphragm, creating a temperature and time dependent unbalance in the measuring bridge. The stress and its impact on sensor performance can be reduced by use of an intermediate layer of a soft polymer between the sensing element and the rigid mold, a so-called glob top.

The aim of this work has been to design piezoresistive sensing structures suitable for measuring the mechanical stress which is introduced in the sensing element by the molding process with and without the use of a glob top material. Two different approaches will be presented: In the first approach the package induced stress has been measured by means of the piezoresistive bridge of an existing pressure sensor. In the second approach stress is measured by use of a modified sensor structure which has been specially designed for measuring the package induced stress.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00146-2

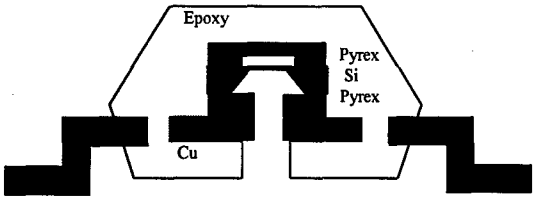



Figure 1 Schematic of pressure sensor placed in molded package

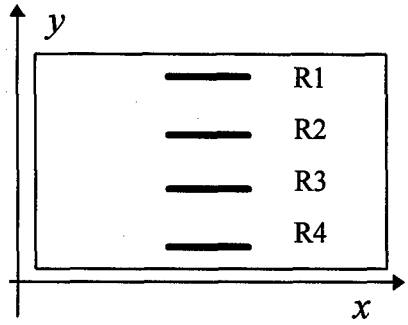



Figure 2 Placement of the resistors $R_1 \ldots R_4$ on the sensor membrane (schematic).

## 2. Sensor functionality

The pressure sensor which has been investigated is an absolute pressure sensor consisting of a silicon chip bonded between two Pyrex glass chips (figure 1). The cavity in the upper glass chip is hermetically sealed and vacuumized, and the lower glass chip has a through-going hole serving as pressure inlet. With increasing external pressure the sensor diaphragm bends upwards, inducing mechanical stress which changes the resistance of four p-type silicon resistors on the diaphragm. Figure 2 shows the schematical placement of the piezoresistors on the membrane. The resistors are interconnected in a Wheatstone bridge. Before transfer molding the sensor chip is attached to a copper lead frame together with an ASIC.

## 3. Package induced stress and sensor drift

The molding material has a thermal expansion coefficient being about 3-4 times as high as that of silicon and Pyrex glass. When temperature is lowered to room temperature after molding, the thermal expansion mismatch induces a compressive stress in the sensor element, creating a bending moment on the diaphragm. The resulting non-uniform stress distribution in the diaphragm affects the four bridge resistors slightly differently, creating a thermal zero shift in the bridge output [1]. With time and under the influence of humidity or thermal excursions the built up stresses may relax or the mechanical properties of the molding material may be altered. This creates a time dependent change in the sensor signal known as long term drift.

By covering the sensor chip with a glob top material with a very low elastic modulus, the sensor chip may be mechanically "decoupled" from the rigid mold. This is expected to reduce the package induced stress in the sensor diaphragm, and hence the thermal zero shift.

### *3.1. Piezoresistive measurement of package stress*

The temperature- and stress-induced relative change of resistance of a piezoresistor defined in the [110] direction of a Si crystal is given by

$$\delta R = \pi_l \sigma_l + \pi_t \sigma_t + \alpha \Delta T \qquad (1)$$

Here $\delta R = (R(\sigma, T_0 + \Delta T) - R_0(0, T_0))/R_0(0, T_0)$ is the relative change of resistance between a stressed resistor $R$ at temperature $T = T_0 + \Delta T$ and a zero stress reference measurement $R_0$ of the same resistor at temperature $T = T_0$ [2]. $\sigma_l$ and $\sigma_t$ are the longitudal and transversal stress components relative to the resistor orientation, and $\alpha$ is the temperature coefficient of resistivity. In (1), the stress component normal to the chip surface ("z stress") is assumed to be zero. The parameters $\pi_l = (\pi_{11} + \pi_{12} + \pi_{44})/2$ and $\pi_t = (\pi_{11} + \pi_{12} - \pi_{44})/2$ are termed the longitudal and the transversal piezoresistive coefficient. The value and temperature dependence of the coefficients $\pi_{11}$, $\pi_{12}$ and $\pi_{44}$, which depend on the doping type, have been determined by Kanda [4]. In this work only p-type piezoresistors have been employed.

By measuring the resistance of a piezoresistor before and after the molding process, the package induced stress in the resistor can be estimated by use of (1). Because the parameter $\alpha$ may vary depending on the processing conditions, it is important to measure the values $R$ and $R_0$ at the same temperature, so that $\Delta T = 0$.

As the value of the piezoresistors depend on both the transversal and the longitudal stress, it is not possible to determine the absolute value of both stress components based on measurements on only one resistor. To measure the two stress components separately, it is necessary to measure the change of resistance for two resistors which are directed along different crystal directions [2].

*3.2 Measurements on existing sensor structure*

As seen by figure 2, the bridge resistors of the existing sensor are all oriented in the $x$ direction, and it is consequently not possible to measure the stress components $\sigma_l$ and $\sigma_t$ separately. However, by measuring the relative change of resistance $\delta R_1 ... \delta R_4$ for these resistors, a rough estimate for the difference $(\sigma_l - \sigma_t)$ between the induced longitudal and transversal stress can be obtained at each of the resistor locations $R_i$. For p type resistors, if $(\sigma_l - \sigma_t)$ is not much smaller than $(\sigma_l + \sigma_t)$, it may be shown that $(\sigma_l - \sigma_t)$ can be approximated by the following expression [1]:

$$(\sigma_l - \sigma_t)_i \approx 2\frac{R_i - R_{i0}}{R_{i0}(\pi_l - \pi_t)} \qquad (2)$$

*3.3 A specially designed sensor structure*

To be able to measure both the $x$ and the $y$ stress on the diaphragm (see figure 2), the resistor layout on the sensor diaphragm was changed, and a batch of modified sensor structures was fabricated in the same process as is used to fabricate the original pressure sensors.

In the modified structure, each of the resistors $R_1...R_4$ is replaced by a rosette consisting of two mutually perpendicular resistors. The layout of the rosette is shown in figure 3.

The rosette geometry is chosen so that most of the resistor length falls within the area of the resistors of the original device, while keeping the length of the two rosette resistors approximately equal.

After measuring the stress induced relative change of resistance for the two resistors in each rosette, the $x$ and $y$ stress can be calculated from the equations:

$$\sigma_x - \sigma_y = \frac{\delta R_x - \delta R_y}{\pi_{44}} \qquad (3a)$$

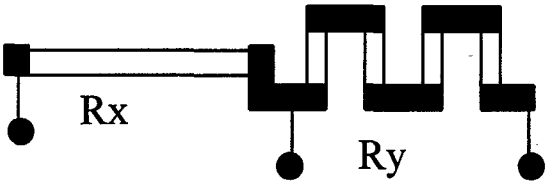



Figure 3 Rosette structure consisting of two mutually perpendicular piezoresistors. The dark areas are higly doped areas serving as low-ohmic contacts between the various areas of the resistor.

$$\sigma_x + \sigma_y = \frac{\delta R_x + \delta R_y}{\pi_{11} + \pi_{12}} \qquad (3b)$$

One of the main advantages of using this two-resistor structure is that the measurement of $(\sigma_x - \sigma_y)$ is inherently compensated for uncertainties in the temperature measurements [2]. Sample-to-sample variations in resistance are also partly corrected for.

For the measurement of $(\sigma_x + \sigma_y)$ there is no such temperature compensation, and a non-zero and unknown $\Delta T$ will introduce large errors in the measurements [2]. To be able to obtain accurate values for $(\sigma_x + \sigma_y)$, it is therefore necessary to measure the temperature of the sample with as high precision as possible. To allow for an accurate determination of the resistor temperature at the time of measurement, a temperature sensing diode was also included on the chip.

## 4. Experimental

A number of pressure sensor chips and specially designed structures were processed and attached to a leadframe. The two types of structures were processed on the same mask set. Before transfer molding, half of the samples were covered with glob top. The molded samples were placed in a temperature chamber, and the resistances $R_1..R_4$ were measured as a function of temperature.

Unfortunately it is not possible to achieve accurate measurements of the bridge resistors before they are molded in the package. The "stress-free" reference resistance $R_0$ must therefore be measured on identical sensor structures that are placed in packages where the stress is known to be close to zero.

For these experiments, the temperature dependence of the stress free reference resistance was determined by performing similar measurements

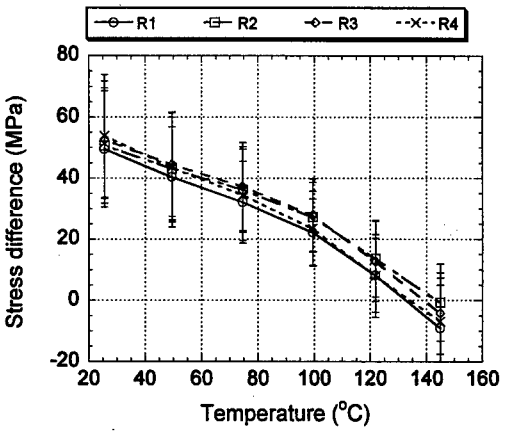



Figure 4 Measurement of ($\sigma_x$ - $\sigma_y$) for samples without glob top. Measurements on existing sensor from a previous batch. (From [1])

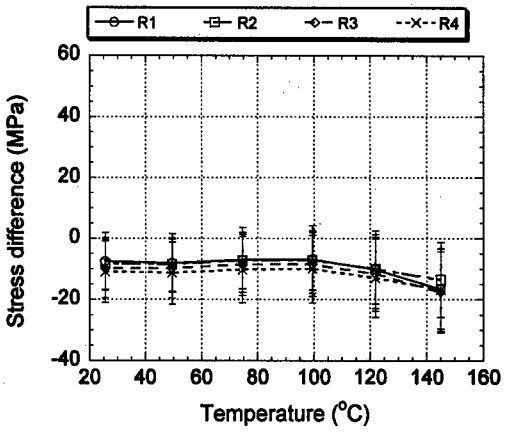



Figure 5 Measurement of ($\sigma_x$ - $\sigma_y$) for samples with glob top. Measurements on existing sensor from a previous batch. (From [1])

on a sensor chip attached to a Pyrex tube in a TO can. Because of the matching thermal expansion of silicon and Pyrex, the package stress here can be assumed to be very low.

Because the zero-stress reference resistance are not measured on the molded samples themselves, statistical variations in doping level or zero-stress resistance will lead to uncertainties in the stress measurements. To avoid the influence of systematic batch-to-batch or wafer-to-wafer process variations, it is important that the reference samples are picked from the same wafer as the molded samples. To reduce the statistical uncertainty it is necessary to perform measurements on as many molded and reference samples as possible. For most of the experiments in this work six samples of each type have been used, and the average value and standard deviation has been calculated in each case.

All measurements were carried out at atmospheric pressure. The effect of the pressure on the sensor output was corrected for by measuring the pressure sensitivity for each of the sensor types.

For the existing sensor, the difference between longitudal and transversal stress was calculated from equation (2). For the specially designed structures, equations (3a) and (3b) were used to calculate ($\sigma_x$ - $\sigma_y$) and ($\sigma_x$ + $\sigma_y$). Before the measurements, the samples were subjected to 50 thermal cycles between -40 and 120 °C.

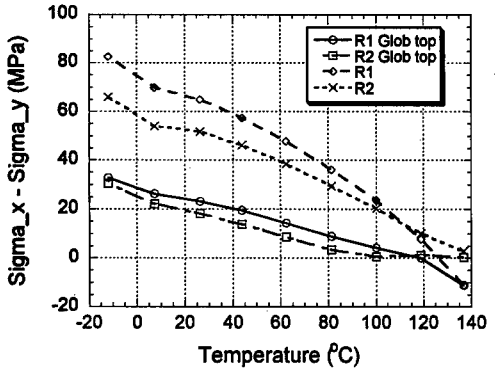



Figure 6 Measurement of ($\sigma_x$ - $\sigma_y$) for samples with and without glob top. Measurements on existing sensor from the new batch.

## 5. Results and discussion

### *5.1 Measurements on existing sensor structure*

Measurements of the induced stress difference ($\sigma_x$ - $\sigma_y$), using the existing sensor structure from a previous batch, have been published in [1]. The obtained results are summarized in figures 4 (samples without glob top) and 5 (samples with glob top). For the samples without glob top, the measurements show that the induced stress difference increases with decreasing temperature, with a zero crossing at around 140 °C. Note that for low temperatures, the gradient of the R1 and R4 curves are somewhat higher than for R2 and R3, leading to a thermal zero shift of the bridge output.

For the samples with glob top, is seen that the stress has been considerably reduced as compared to the samples without glob top.

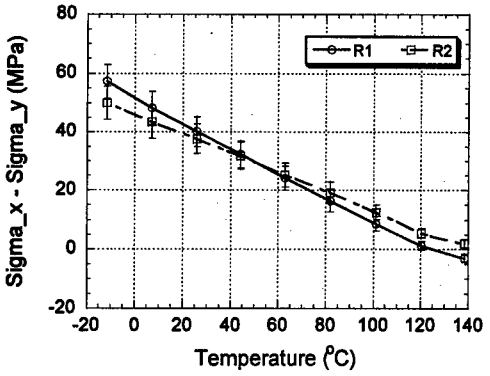



Figure 7 Measurement of $(\sigma_x - \sigma_y)$ for R1 and R2 for samples without glob top. Measurements on modified sensor structure. Standard deviation is indicated by the error bars.

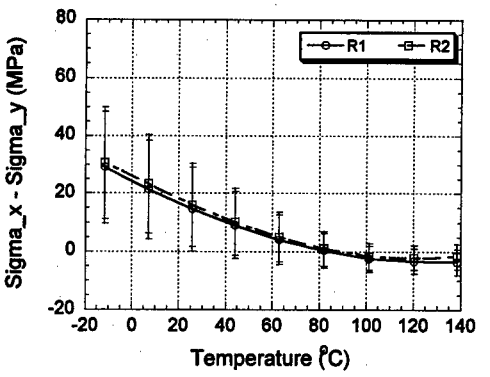



Figure 8 Measurement of $(\sigma_x - \sigma_y)$ for R1 and R2 for samples with glob top. Measurements on modified sensor structure. Standard deviation is indicated by the error bars.

Figure 6 shows corresponding measurements of $(\sigma_x - \sigma_y)$ using samples from the new batch. The measurements in figure 6 are based on the average of two samples, and have only carried out for the that resistors R1 and R2.

Comparing the results from the two batches, it is seen that the stresses in figure 6 are somewhat higher, especially for the samples with glob top. No explanation has been found for this difference, but it may perhaps be due to batch-to-batch variations in the encapsulation process. It is also possible that the results in figure 4 have systematical errors because the reference samples were here taken from a different batch [1].

*5.2 Measurements on specially designed sensor structure*

Figure 7 and 8 show measurements of $(\sigma_x - \sigma_y)$ for samples without and with glob top. As seen by the curves, the measurements of are in relatively good agreement with the results obtained with the existing sensor structure from the same batch (figure 6).

For the samples without glob top it is seen that the standard deviation of the measurements has been considerably reduced as compared to the results in figure 4. This shows the advantage of using a resistor structure which compensates for uncertainties in temperature as well as sample-to-sample variations in the resistance. For the samples with glob top the larger variation at low temperatures may reflect a larger spread in the induced stress.

FEM simulations of $(\sigma_x - \sigma_y)$ at 25°C (samples without glob top) indicate a thermal gradient of -0.9 MPa/°C for R1 and -0.7 MPa/°C for R2 [5]. This is around twice as high as the measured values (figure 7).

FEM simulations of glob-topped samples have not been carried out for this process, but simulations on comparable structures indicate that the absolute value of $(\sigma_x - \sigma_y)$ at 25°C is below 10 MPa [5]. The corresponding measurement is $20 \pm 20$ MPa (figure 8).

Figure 9 and 10 show measurements of $(\sigma_x + \sigma_y)$ for samples without and with glob top. It is seen that for both types of samples, there is a measured compressive stress in the silicon diaphragm. The standard deviation of the measurements is however very large. The large uncertainty probably arises from a combination of incomplete temperature compensation and the very low sensitivity of this measurement when p type resistors are employed [2]. Sample-to-sample variations in resistance will also give rise to large uncertainties.

To improve the measurement of $(\sigma_x + \sigma_y)$, it is probably necessary to measure a much larger amount of samples, or to perform the stress reference measurements on individual unpackaged samples before carrying out the molding process.

FEM simulations of $(\sigma_x + \sigma_y)$ indicate a room-temperature thermal gradient of around 2 MPa/°C for both R1 and R2 (samples without glob top). For samples with glob top the room temerature absolute value of $(\sigma_x + \sigma_y)$ is calculated to around -30 MPa.

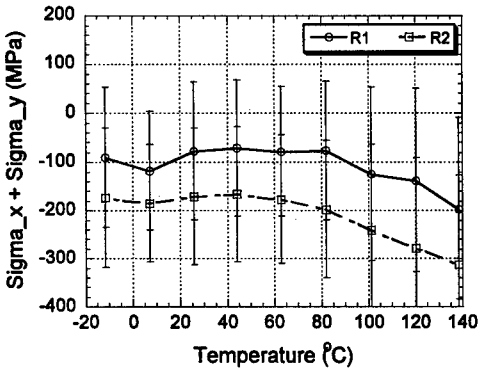



Figure 9 Measurement of ($\sigma_x$ + $\sigma_y$) for R1 and R2 for samples without glob top. Measurements on modified sensor structure. Standard deviation is indicated by the error bars.

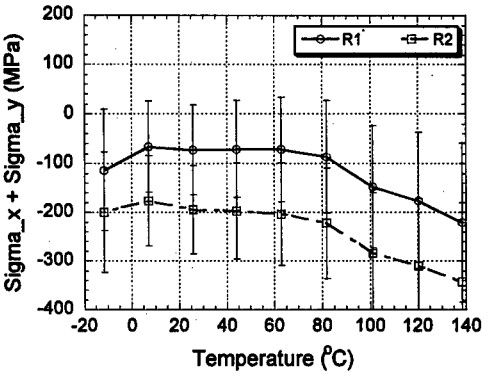



Figure 10 Measurement of ($\sigma_x$ + $\sigma_y$) for R1 and R2 for samples with glob top. Measurements on modified sensor structure. Standard deviation is indicated by the error bars.

## 6. Conclusion

The induced stress in a pressure sensor diaphragm has been measured piezoresistively by use of a specially designed sensor structure as well as an existing pressure sensor. The difference ($\sigma_x$ - $\sigma_y$) between x and y normal stress can be measured by both types of structures, but the specially designed structure gives advantages such as automatic temperature compensation and lower standard deviation. For both structures, errors are induced in the stress measurements because the zero-stress reference resistance cannot be measured for individual sensors.

For transfer molded devices without glob top, the average value of ($\sigma_x$ - $\sigma_y$) at room temperature has been measured to around 50 MPa. By use of a glob top material the induced stress is reduced to around 20 MPa.

For the specially designed structure the magnitude of ($\sigma_x$ + $\sigma_y$) has also been measured, and the results indicate high levels of compressive stress in samples both with and without glob top. The uncertainty of these measurements is however very high, and a larger statistical sample is needed to determine the magnitude of the induced stress.

FEM simulations of the thermal gradient of ($\sigma_x$ - $\sigma_y$) are around a a factor of two higher than the measured values for samples without glob top.

## Acknowledgements

The authors wish to thank Gjermund Kittilsland, Henning Sørensen, Roy Grelland and the rest of the staff at SensoNor who have been helpful in design or processing of the experimental samples.

## References

[1] J. B. Nysæther, A. Larsen, B. Liverød and P. Ohlckers, "Measurement of Package Induced Stress and Thermal Zero Shift in Transfer Molded Silicon Piezoresistive Pressure Sensors", Proceedings of Micromechanics Europe 1997, Southampton, UK. To be published in Journal of Micromechanics and Microengineeering

[2] Jaeger, R. C., Suhling, J. C., and Ramani, R., "Errors Associated with the Design, Calibration and Application of Piezoresistive Stress Sensors in (100) Silicon", IEEE Transactions on Component, Packaging and Manufacturing Technology, Part B, Vol 17, No 1, 1994

[3] G. Kelly, Packaging Induced Thermo-mechanical Stress, Proceedings of Micromechanics Europe 1996, Barcelona, Spain, p 297-299.

[4] Y. Kanda, "Piezoresistance effect of silicon" Sensors and Actuators A, 28, pp 83-91, 1991

[5] Results from SensoNor

PERGAMON

Microelectronics Reliability 38 (1998) 1277–1286

MICROELECTRONICS RELIABILITY

# Electronic systems packaging: future reliability challenges

J. Barrett

*National Microelectronics Research Centre, University College, Cork, Ireland*

**Abstract**

Packaging has a dominant effect on electronic system cost, performance, weight, size and long term reliability. Recent years have therefore seen rapid developments in packaging to meet the challenges of miniaturisation and cost reduction while delivering increased electrical performance and reliability. Meeting the reliability demands posed by these developments already presents major challenges to the reliability practitioner. However, the technology roadmaps for integrated circuits and electronic systems imply a continuous increase in the scale of existing challenges while emerging component and system technologies, such as optical interconnections and microsystems, will present completely new challenges. These emerging challenges will mean that the reliability practitioner will increasingly be involved from the very first phases of component or system conception and design all the way through to production and "green" disposal. Further, to ensure that reliability targets are realistically set and met, the reliability practitioner will increasingly function as part of a multidisciplinary team, many of whose members may not belong to the traditional electronics disciplines. By examining current reliability challenges and by the use of technology roadmaps, this paper tries to forecast future reliability challenges in electronic packaging. In the context of this paper the term "Packaging" is used to encompass the various assembly and interconnection technologies and techniques which are used to build an electronic component or system. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

The electronic system packaging reliability challenge covers a very broad spectrum: at one extreme there are short-time-to-profit, short product life cycle, cost driven volume products, such as mobile phones and personal computers, which operate in relatively benign environments; at the other extreme are critical systems, such as those for avionics applications, which may be many years in operation in harsh environments and are characterised by longer development cycles, smaller production volumes and higher costs. At each extreme and across the spectrum as a whole the packaging reliability practitioner has a role in ensuring a level of system reliability which meets the requirements of the end user and the end use environment. Given the diversity of materials and components which are used in electronic system manufacture and the complexity of many of the individual system components used, particularly integrated circuits, there are potentially many ways in which an electronic system may fail. The packaging reliability practitioner has by necessity been, or has had to become, an electronic "jack-of-all-trades" requiring expertise in materials, ICs, passive components, product engineering, manufacturing and the environment as well as in

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00129-2

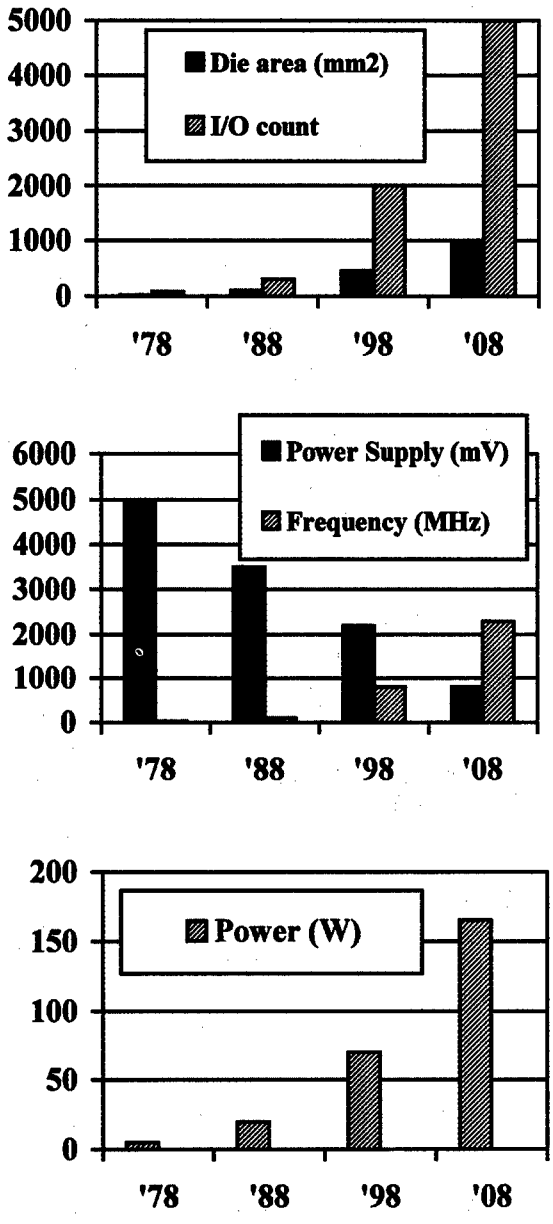
the more obvious disciplines of reliability physics and statistics. Increasingly, packaging reliability has become a team, rather than an individual, discipline and it is a discipline which will become increasingly challenging, driven by developments such as:

- The growth in IC complexity
- Portability
- Harsh environment applications
- Opto-packaging and interconnection
- Green electronics
- Microsystems and micromechanical systems

These developments and their implications for future systems packaging reliability are discussed in the remainder of this paper.

## 2. Growth in IC complexity

Moore's Law, which predicts a doubling of IC processing power every 18 months has proved to be more or less valid since it was first proposed by Gordon Moore, one of the founders of Intel, in 1965. This has been achieved through IC feature size reductions and through increased fabrication yields which allow larger die sizes. Inevitably this has set equally complex challenges for packaging engineers as leading edge IC area, operating frequency and power dissipation have increased, as shown in Figures 1.(a) -.(c) [1]. These evolutions have been particularly rapid in the last five years and the pace shows no sign of slowing in the next ten years. At the same time, the predicted cost per I/O of high pincount packages requires a 50% cost reduction in the next 10 years [1], a trend which is typical of cost evolution in the past decade. The packaging engineer has had and will have the task of providing more complex packaging solutions at reduced cost. The technical challenges facing the reliability practitioner have grown and will continue to grow accordingly. A review of these challenges in the electrical, thermal and thermomechanical domains follows.




Figures 1.(a)-(c). Evolution of IC area and I/O count; operating frequency and power supply voltage; and power dissipation vs. time.

### *2.1. Electrical reliability*

Electrical reliability involves the integrity of signal propagation in a system under all operating conditions. Factors which are making the reliability challenge more difficult include:

- higher operating frequencies
- lower power supply voltages
- miniaturisation
- EMC/RFI regulatory measures

Higher operating frequencies and lower supply voltages make the task of ensuring signal propagation integrity increasingly difficult. As digital frequencies enter the GHz region, RF and microwave parameters, which can largely be ignored at lower frequencies, must be taken into account and each individual signal trace in a package or circuit board must be designed with impedance and noise control in mind. In this context, a traditional wire-bond will not be suitable for chip-to-board interconnection as its impedance cannot be tightly controlled. Miniaturisation simplifies this challenge to a certain extent, in that signal traces are overall made shorter, reducing their parasitic loading and coupling lengths.

However, miniaturisation does not in general imply a reduction in package or system complexity: the pressure, largely produced by the desire for portability, continuously exists to pack more functionality into smaller outlines. This is usually achieved by a combination of greater integration at the IC level and by increasing the IC packing density at board level. Packing densities (ratio of total area of all ICs to the area of the circuit board) of 10% were relatively recently regarded as high; it is now not unusual to see packing densities greater than 50% in high density multichip modules (see Figure 2).

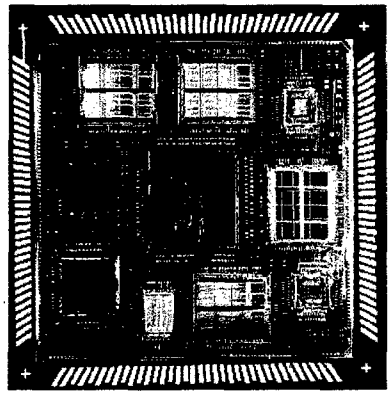


Figure 2. A Pentium silicon-on-silicon multichip module with high packing density (*photo courtesy Europractice-MCM/ETH*).

This level of packing is achieved by the use of fine pitch wiring boards where track dimensions and spaces are driven to the minimum allowed by manufacturing technology. Consequent increases in self-inductance and in electrical coupling between adjacent lines results.

Further complicating the task are the tighter regulations on electromagnetic compatibility and radio frequency interference (EMC/RFI). Design tools for EMC/RFI have lagged behind the needs of increasingly complex, miniaturised systems operating at high frequencies. EMC/RFI precautions in many system designs are still rooted in rules-of-thumb and designer experience rather than in analysis of sources and effects and their impact on electrical reliability. Designers still have to wait for testing results from a compliance laboratory before having confidence in EMC/RFI system performance. Solving this issue will be one of the major challenges in design for electrical reliability.

Current approaches to the electrical reliability issue include:

- Minimising chip-to-board interconnection parasitics by almost complete elimination of intermediate packaging through the use of flip-chip assembly. This also eases the thermal reliability task but significantly increases the challenge of thermomechanical reliability (see section 2.2).

- Use of new lower dielectric constant organic materials and controlled impedance interconnections in packages and circuit boards. However, long term interfacial stability of these dielectric materials under thermal, thermomechanical and environmental stress then enters the reliability challenge.

- Development of more sophisticated integrated design tools which can efficiently analyse all aspects of electrical reliability. This is an important requirement on the electrical reliability roadmap.

In the future, it is likely that optical interconnection will need to be used at board level for critical signal paths (see Section 5)

### *2.2. Thermal and thermomechanical reliability*

The trend towards increasing density in interconnections is matched by increased power

density in systems as the forces of increased operating frequency and increased packing density combine to increase IC power dissipation and the number of ICs in a given board area. Meeting the challenge of heat removal is one the most important aspects of package and system design and increasingly requires the use of new materials (e.g. thermal interface pads and phase change materials) and techniques (e.g. integrated micro-channel heat pipes, micro-fans and heatsinks). Overall, however, the pressure exists to allow higher junction temperatures and consequent exposure of packaging materials to higher continuous temperatures. This is combined with the increased use of organic materials to reduce electrical parasitics, cost and weight. The reliability challenge therefore becomes that of ensuring stability of the organic materials themselves, and their interfaces to other organic and inorganic materials, under continuous exposure to more elevated temperatures. To quote from the SIA roadmap for semiconductors [1]:

"Standard methods and acceptance criteria for interfacial adhesion are lacking. Fundamental work is needed to establish adhesion strength and degradation rate versus environmental factors, all as a function of interfacial physical and chemical properties. New materials are being added to an already complex, poorly understood interfacial adhesion reliability issue."

Adding to the thermal stress on materials and interfaces is the problem of thermomechanical stress. Increases in die area, coupled with higher temperature gradients and transients, and the additional complication of the use of organic materials whose coefficients of thermal expansion (CTE) are usually higher than that of silicon, mean that both absolute and cyclic thermomechanical stresses on ICs, components and materials will continue to increase. This is without even considering the effect of cycles in the external thermal environment. Compressive, tensile and shear forces exist at every level of packaging and can lead to delamination, fatigue and cracking of ICs, plastics and solder joints.

*2.2.1. Flip-chip assembly*

The problem of thermomechanical stress is becoming particularly acute in flip-chip assembly of ICs. In flip-chip, metal bumps are deposited directly on the bond-pads of the IC and the IC is then directly reflow soldered to its circuit board, as shown in Figure 3. This almost completely eliminates the electrical parasitics associated with the chip-to-board interconnections. Flip-chip also allows an almost unlimited number of I/O connections as it facilitates area-array distribution of connections across the entire IC surface. However, the IC is now connected with less than a 100 μm gap to its substrate. In the early application of flip-chip, ceramic substrates were used by IBM [2] which had a reasonable CTE match to silicon and IC areas were considerably smaller than current and projected sizes. Even today, ceramic substrates are used for the Motorola PowerPC chip [3]. However, for reasons already explained above, organic substrates are increasingly used for flip-chip interconnection. The CTEs of even the most optimised organic substrates remain sufficiently higher than that of silicon to cause early fatigue failure of the micro-solder joints (<.001 $mm^3$ in volume) induced by cyclic thermomechanical stress. The use of organic underfill adhesives [4] has enabled this problem to be alleviated for current generation flip-chip assemblies. However, given projected IC size, power dissipation and I/O, considerable work must be done on materials, interfaces [5], solder metallurgy and thermo-mechanical simulation tools if future generation flip-chip assemblies are to meet their reliability targets.

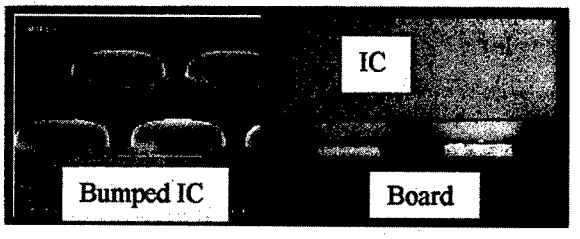



Figure 3. Micrograph of bumped IC and cross-section of flip-chip assembly to circuit board.

*2.2.2 Plastic packaging*

While flip-chip assembly is likely to be the technique of choice for chip-to-board interconnection of leading edge ICs, "conventional" transfer moulded plastic encapsulated packaging will continue to dominate the market in terms of volume. Plastic packaging has made considerable progress since the early stages of dual-in-lines with low reliability caused its use to be restricted in

many hi-rel applications, a restriction which has only in recent years begun to be lifted. Intervening years have seen the emergence of new body formats, progression to ever higher pincounts and improvements in reliability which now allow commercial-off-the-shelf (COTS) plastic packaged ICs to be used even in hi-rel applications. This reliability improvement has been brought about through a combination of materials and process engineering to reduce stress and moisture induced failures, and application of finite element simulation tools to the thermomechanical stress problem. However, the reliability challenges can be expected to continue growing in the coming years in topics such as:

- Stress and moisture issues for large die in very thin packages where there may be only a few hundred microns of plastic over the IC.
- The demand for increased power handling.
- The use of plastic packages in increasingly harsh application environments (see Section 4).
- The ability to handle higher I/O counts in smaller footprints.

Among the tools that will be needed to succeed in meeting these challenges, and those of flip-chip, will be:

- The ability to quickly develop application specific encapsulant materials through the use of techniques such as molecular modelling and synthesis

- Integrated design and simulation tools that will be able to perform "virtual prototyping" of packages through integration of thermal and thermomechanical simulations with reliability models. Essential here also will be validated compact models (and the means to obtain them) for material and interface behaviour in the intended use environment. While it unlikely that absolute accuracy can be obtained from such tools, they should at least be able to quickly assess the comparative reliability performance of different designs and material sets.

## 3. Portability

It is estimated that in 2000, over 50% of all electronic products manufactured will be portable. This will include increasingly sophisticated smart cards as well as the communication, computing and entertainment systems already in use and which are converging to one integrated portable system. New applications are also emerging in portable chemical and biological analysis systems (see Section 7 below). Portability's main system criteria are:

- Light weight and small size
- Low to moderate cost
- Maximum functionality per $cm^3$
- Low power consumption
- Resistance to shock, vibration and water

These are driving many of the issues already discussed in section 3 above:

- Maximum integration on silicon with consequent growth in die size, I/O count and individual die power consumption
- Reduced overall system power consumption
- Use of organic materials
- Use of low profile, thin packaging techniques
- Use of high density circuit boards

To reduce size and increase reliability there is active ongoing research in the integration of passive components onto and into both inorganic and organic circuit boards. Doing this provides more board surface area for ICs (thus increasing the packing density and associated reliability problems) and also eliminates the solder joints associated with discrete components, of which there may be many hundreds even in a relatively small system such as a mobile phone. As solder joints are viewed as weak links in electronic systems, this should, theoretically, lead to increased reliability. However, the circuit board itself will become a more complex materials system, incorporating resistive, dielectric and/or magnetic materials. New challenges in electrical reliability and in materials/interfacial reliability will therefore have to be addressed.

Smart cards, which are among the highest production volume consumer electronic products, present unique problems in terms of packaging reliability. They are in general constrained to credit card thickness and the protection that can be afforded to the IC is minimal. Future smart cards will be expected to have increased functionality while continuing to be able to withstand the wear and tear of daily use. Considerable ingenuity will

be required to develop packaging with adequate reliability and yet be compatible with high volume, very low cost production.

## 4. Harsh environment applications

It has generally proven possible to obtain adequate electronic packaging reliability even in the harshest aerospace and maritime environments. However, this reliability has typically not come cheaply and product development times were usually long and production volumes relatively small. A number of harsh application environments are, however, imposing a reversal of this trend. Dominant among these is automotive electronics but the demand for lower cost communications satellites, commercial aircraft control systems and drill-head electronic systems are also influential. The under-hood automotive environment is perhaps the harshest "conventional" environment in which an electronic package has to function. However, the automotive industry demands high volume production, low unit cost and relatively short system development times. As the sophistication of engine monitoring and control systems increases, electronics is claiming a higher proportion of overall vehicle cost and the trend towards cost reduction in packaging can therefore be predicted to increase. Reliability specifications are however likely to become more stringent as automotive electronics increasingly take over safety critical functions. The reliability challenge set out is therefore to provide high reliability at low cost in a very harsh environment. To meet this challenge, all levels of system reliability will have to be addressed both on and off-chip and sophisticated design for reliability methodologies will have to be developed and applied which will concurrently address many of the issues which are discussed in the other sections of this paper. Successfully meeting this challenge will however provide materials, processes and design methodologies which will support lower cost application of electronic systems in other harsh environments.

## 5. Opto-packaging and interconnection

Due to increasing operating frequencies, optical interconnections, already in use at backplane level, are increasingly likely to be applied at individual board level. One of the first applications which will drive this development is likely to be clock distribution but it also will be increasingly required for critical signal interconnections. Two developments are needed to enable board level optical interconnections: low cost polymeric optical interconnects and a means of allowing conventional silicon ICs to efficiently emit light.

R&D is already underway on optical interconnects that can be integrated into mainstream board manufacturing processes [6]. It is therefore likely that the integrated optical interconnections will at least be available although, in a similar manner to integrated passives, new reliability challenges in materials and interfacial stability will have to be addressed. There does not seem to be an equal possibility that conventional silicon ICs will be able to efficiently emit light. However, low cost, small sized compound semiconductor lasers and LED's do exist and hybridisation techniques can be used to attach these to silicon ICs to allow them to emit light [7]. Important among the reliability issues which need to be addressed include the impact of the pick and place operation on optical device reliability and the effect of cyclic thermomechanical stress on the reliability of the VCSEL itself and its attachment. The device attachment is however only the first step in building the optical hybrid system. The issue of interfacing the VCSEL to the optical interconnection and ensuring the long term mechanical, thermal and environmental integrity of that interface must also be addressed. Add to this that projected complex systems may eventually require the mounting of hundreds of such devices and the scale of the reliability challenge becomes apparent. Package and board level optical interconnection will require, and will be a vehicle to demonstrate, many new developments in reliability science.

## 6. Green electronics

A market and consumer based driver for advances in packaging is that of environmentally friendly electronics which includes both manufacturing and ultimate disposal/recycling. The principle of extended producer responsibility where, for example, an IC packaging company must take ultimate end of life responsibility for its products, has long term technological implications

for the IC packaging process. A common example is that of transfer moulded plastic packages, fabricated in volumes of billions per year. The plastic material used is largely thermoset epoxy with a high silica filler content and other minor constituents (including environmentally unfriendly antimony and bromium based flame retardant compounds). Once the epoxy is cured it becomes largely impervious to chemical attack, with only boiling fuming nitric acid (itself now a restricted material) capable of chemically stripping the epoxy so that the chip and leadframe can be recovered and recycled. Although organic "digesters" are under development for the epoxy, the principal technique for epoxy removal at end of life is incineration, with release of undesirable by-products. To facilitate recycling of plastic packages, a pressure is being exerted to use thermoplastic encapsulants, a move that will raise many new reliability challenges.

The environmental issue which has received the most publicity is however that of lead-based (Pb) solders. It has been threatened for at least the last five years that lead would be legislatively banned from electronic products in Europe and elsewhere. That threat has failed to materialise yet many companies (e.g. Nortel and Hitachi) have announced the removal of lead from their electronic products. There is also active European research into the development of lead free solder alloys compatible with current electronics assembly processes. The driver for these developments is largely driven by perceived market demand, fuelled by consumer environmental concern, for lead free products. From a reliability point of view, there exist decades of accumulated reliability data for tin-lead solders and reliability predictions have become increasingly accurate. With the introduction of lead free alloys, urgent work has had to be carried out and will continue to be required to evaluate the reliability of lead-free alloys in all their applications from large power device attach down to flip-chip solder joints.

Plastic IC packaging and solder alloys are only two examples - most printed circuit boards are thermoset epoxy based and contain copper, nickel, cadmium, lead and gold which are undesirable pollutants if allowed into drinking water. Recycling of PCBs will demand reclamation of these metals from both outer and buried layers and a more easily removed material than thermoset epoxy will be needed. Table 50, which is too large to reproduce here, of the SIA roadmap for semiconductors [1], lists further examples of material and manufacturing changes that will be needed to meet environmental requirements. The consumer can hopefully look forward to a cleaner future; reliability practitioners can certainly look forward to a challenging one where there will be a demand for rapid assessment of the reliability implications of using alternative material sets. It is unlikely that there will be time available to do extensive reliability testing of many different alternatives. The tools however already do exist, and will become more sophisticated, to simulate the electrical, thermal and thermomechanical effects of using alternative materials. What is not readily available is a fast and accurate methodology for obtaining material and interfacial properties of materials in configurations which are representative of their end use application. For example, if a thermoplastic is going to be used as a 25 μm thick layer, we will need to know how it behaves electrically, thermally and mechanically at that thickness and not as a bulk sample. This information will allow the most promising material sets to be selected. These can then subjected to the reliability testing which is unlikely, at least in the near term, to be replaced by reliability simulators for complex multi-material and multi-component systems with multiple potential failure mechanisms. The development of methodologies for efficient measurement of material properties, and how these properties change with temperature, humidity etc., is an important challenge.

## 7. Microsystems

Microsystems technology (MST), defined as the integration of sensors, actuators and microelectronics in a monolithic or hybrid circuit, and its counterpart micro-electromechanical systems (MEMS) are regarded as being at a similar technology and market stage as microelectronics 30 years ago [13]. Relatively simple devices, such as automotive airbag sensor/actuator ICs, are being used in high volumes while increasingly complex systems are being demonstrated in many laboratories and are beginning to appear on the market. Examples include:

- Micro-total analysis systems (μTAS) for chemical/biological analysis which contain pumps, chemical or biological sensors and signal processing on a single chip or hybrid.
- Micro-metrology systems such as miniaturised spectrometers and gyroscopes.
- Micro-optical systems containing integrated lasers, adjustable lenses, mirrors and filters such as the Texas Instruments micro-mirror projection IC [8].

Many other examples exist but two primary barriers to volume commercialisation are consistently mentioned: cost efficient packaging and reliability. The reasons for this can be readily seen if we examine some of the new reliability challenges which MST/MEMS present.

### *7.1. Low or zero stress packaging*

Many silicon based MST/MEMS components contain micro-machined membranes or beams, used for sensing and/or actuation, which are of the order of a few microns thick but which may be hundreds or thousands of microns in length/width. These structures are very sensitive to mechanical stress which may cause incorrect operation or complete failure. At the same time, these components must undergo the "normal" packaging processes of die attach, electrical connection and encapsulation, all of which impose bending stresses of various sorts on the component, especially as the preferred route for low cost is to use organic packaging materials. Figure 5.(a) shows a stress simulation of one corner of a micro-machined membrane subjected to plastic encapsulation stresses. Critical stresses were predicted to exist in the membrane (lower left) which exceeded the membrane strength and in a micro-machined channel (lower right). Failure analysis confirmed the simulation and the cracked membrane is shown, after decapsulation, in Figure 5.(b) [9]. Methodologies must therefore be found which mechanically and thermomechanically decouple the micro-machined structures from the package without compromising reliability.

### *7.2. Particulate and contaminant free packaging*

In MST/MEMS components which contain micro-mechanical moving parts of the order of tens or hundreds of microns in size, the presence of particulates or contaminants can compromise system operation and reliability. Indeed, wear in moving micro-components can itself generate particulates even if they are absent from the package as fabricated [13]. In components which use biological layers for sensing, biological contamination from packaging materials and the effect of packaging thermal processes on the bio-layers must be taken into account. A micro-sensor which is intended to measure heavy metal (Pb, Cu, Zn, Cd) content in water to an accuracy of parts per billion cannot be contaminated by leaching of such metals from packaging materials [10].

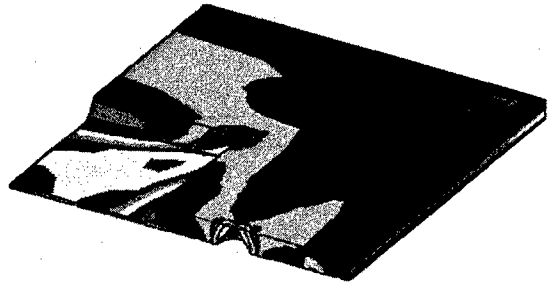


Figure 5.(a). Thermomechanical stress simulation of a micro-machined silicon structure

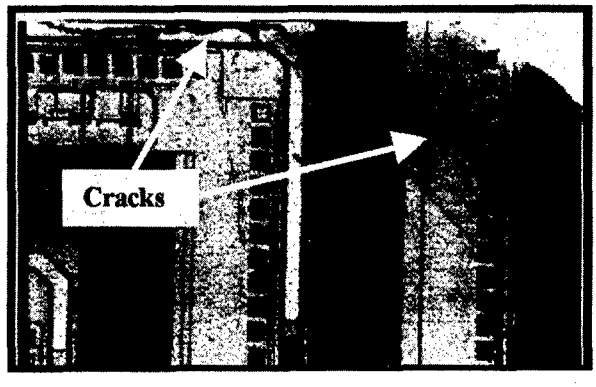



Figure 5.(b). Stress induced cracks in the micro-machined silicon structure of Figure 5.(a).

### 7.3. "Multimedia" input-output

The problem of dealing with electrical input-outputs was discussed in Section 2.1 above and the issue of optical input-output was referred to in Section 6. If added to this is fluidic and/or gaseous input-output, the challenge is considerably complicated. At the simplest level is a sensor IC which has to be immersed in ambient temperature aqueous analytes and even this represents a considerable challenge in defining and proving the reliability of a low cost packaging process which exposes the sensing element while protecting the microelectronics [10]. The extension of such a packaging process to higher temperature or more chemically aggressive analytes is a considerable reliability challenge. However, many MST/MEMS components are being developed which will include active flow-through systems driven by integrated micro-pumps. Such components require "micro-plumbing" input-output and internal "packaging" to isolate the fluid from microelectronics sections of the component. Some of these components will also use micro-spectrometers to perform analysis of the passing fluid and an optical interface will therefore also have to be provided. Add the complications that such a system should be compact, portable, low-cost and recyclable and the scale of the reliability challenge becomes apparent.

MST/MEMS components present most of the reliability challenges associated with conventional microelectronics and add many new ones. A good example of the additional challenges to be considered, and discovered, in bringing a MEMS component to market is given in the IRPS'98 paper on the reliability engineering of the Texas Instruments micro-mirror projection system [8].

## 8. Review and summary

This paper has attempted to forecast the major reliability challenges that will need to be faced in the component and systems packaging fields in the coming years. It is obviously difficult to discuss all the possible challenges as it is similarly difficult to be quantitatively specific on their nature or timing; however some overall trends can be forecast:

- Reliability is moving and will continue to move further back in the system development cycle with design for reliability becoming an integral part of the cycle from the beginning. Increasingly, packaging and reliability will be application driven and system focused. The traditional sequential, and not necessarily interactive, process of chip design → chip-fab → package → board → box → burn-in/reliability test will not be adequate. Concurrent multi-disciplinary effort will be required to deliver a reliable system to time and budget.

- There is a growing expectation of increased reliability at reduced cost and of shorter system development cycles. The time and budget for prototyping and testing will therefore progressively shrink. Simulation and virtual prototyping will therefore increasingly become important reliability activities. They will however only be viable if the inputs on material data and behaviour are accurate.

- Reliability challenges in "conventional" microelectronics and opto-electronics packaging can be reasonably forecast but the pace of progress will be very fast over the next ten years. The reliability community therefore needs to begin work now on developing its own roadmaps to meet these challenges. The recently published European packaging roadmap is an example [11] although it is not specifically focused on reliability issues.

- New challenges are likely to continue appearing from sources such as environmental legislation and consumer concern. Methodologies to quickly assess the reliability impact of these and to recommend optimum solutions will be required.

- MST and MEMS technology offers probably the greatest new challenge in packaging reliability. New materials will be used and many extra material parameters and behaviours will need to be considered in designing for and assessing reliability. The multidisciplinary concurrent development teams that are important for microelectronic systems will be essential for MST and MEMS packaging. As well as consisting of the traditional electronic, physics, materials and mechanical disciplines which, to at least some extent share a common

technical language, the teams will also contain disciplines such as micro-chemistry and micro-biology. The integration of these disciplines to produce a reliable product will itself be an interesting challenge. To quote the IEEE Transactions on CPMT: "Electronic Packaging: Cross-Discipline is the Only Discipline" [12].

**Acknowledgements**

The content of this paper has been informed by discussions with many colleagues in NMRC, with partners in collaborative projects and at conferences. Too many to mention individually, their collective enthusiasm to discuss these issues is gratefully acknowledged.

**References**

[1] Semiconductor Industry Association, The National Technology Roadmap for Semiconductors, 1997. (http://notes.sematech.org/97pelec.htm)

[2] Totta P.A. Flip-chip Solder Terminals. 21st Electronics Components Conference (1971) 275.

[3] Parry J., Rosten H. and Kromann G.B. The development of component-level thermal compact models of a C4/CBGA interconnect technology: the Motorola PowerPC 603 and PowerPC 604 RISC microprocessors". IEEE Trans. Components, Hybrids and Manufacturing Technol. Part A Vol. 21 No. 1 (March 1998) 104-112.

[4] Giesler J, O'Malley G, Williams M and Machuga S. Flip-chip on board connection technology: process characterisation and reliability. IEEE Trans. Components, Hybrids and Manufacturing Technol. vol 17 (Aug. 1994) 256-262.

[5] Dai X , Brillhart M V. and Ho P S. Polymer interfacial Adhesion in microelectronic assemblies. Proc. 48th Electronics Components and Technology Conf. (1998) 132-137.

[6] Dhoedt B , Baets R, Van Daele P , Heremans P, Van Campenhout J, Hall J , Michalzik R , Schmid A, Thienpont H, Vounckx R , Neyer A, O'Brien D C, Van Koetsem J. Optically interconnected integrated circuits to solve the CMOS interconnect bottleneck, Proc. 48th Electronics Components and Technology Conf. (1998) 992-998.

[7] Liu YS, Wojnarowski RJ, Hennessy WA, Piacente PA, Rowlette J, Kadar-Kallen M, Stack J, Liu , Peczalski A , Nahata A and Yardley J. Plastic VCSEL array packaging and high density polymer waveguides for board and vackplane optical interconnect, Proc. 48th Electronics Components and Technology Conf. (1998) 999-1005.

[8] Douglas M R. Lifetime estimates and unique failure mechanisms of the digital micromirror device. Proc. of the Int. Rel. Phys. Symp. (1998) 1-8.

[9] Kelly G, Alderman J, Lyden C, Barrett J, Val A, Camon H, and Sbbia Z. 3D packaging solutions for a silicon micro-pump, Proc. 47th Electronic Components and Technology Conference (1997) 1227-1234.

[10] Barrett J, Garde A, Lane B, Stam F. A reliable, low cost packaging and assembly process for integrated environmental sensors, Proc. of the European Workshop on VLSI and Microsystem Packaging Techniques and Manufacturing Technologies (1998).

[11] The Netpack European Packaging Roadmap (1998).

[12] Pecht M G. Electronic packaging: cross-discipline is the only discipline. IEEE Trans. Components, Hybrids and Manufacturing Technol. Part A Vol. 21 No. 1 (March 1998) 192-194.

[13] Miller S L, Rodgers M S, LaVigne G, Sniegowski J J, Clews P, Tanner D M, Peterson K A, "Failure modes in surface micromachined micro-electromechanical actuators. Proc. of the Int. Rel. Phys. Symp. (1998) 17-25.

PERGAMON

Microelectronics Reliability 38 (1998) 1287–1291

MICROELECTRONICS RELIABILITY

# A concept to relate wire bonding parameters to bondability and ball bond reliability

Z. N. Liang[a], F. G. Kuper[a], and M. S. Chen[b]

[a] *Philips Semiconductors, MOS4YOU, Gerstweg 2, 6534 AE Nijmegen, The Netherlands*
[b] *Philips Semiconductors, PSK, 10 Chin 5th road, N.E.P.Z., P. O. Box 35-48, Kaohsiung, Taiwan, R. O. China*

**Abstract**

The effects of wire bonding parameters on bondability and ball bond reliability have been investigated. Bondability is characterized by ball shear stress (ball shear force per unit area) and ball bond reliability by median time to failure during in-situ ball bond degradation measurements. By introducing the concept of a reduced bonding parameter (RBP), a combination of all bonding parameters, we are able to relate the bonding parameters to bondability and ball bond reliability. With the appropriate RBP, ball shear force, ball shear stress, and ball bond reliability appear to be well-behaved functions of the RBP for a wide range of settings. This provides us with simple analytical tool for optimizing bonding parameter windows.
© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

The formation of voids in Au-Al intermetallic compounds degrades the long-term reliability of Au wire bonds to Al bond pads [1,2,3]. It is evident that bonding parameters are influencing factors not only on ball bond strength but also on void formation [1,4]. In order to improve ball bond reliability, a lot of effort has been spent on optimizing the bonding parameters [5,6,7,8] and to reduce wire bond process variation [9].

In this paper, we focus on three bonding parameters: ultrasonic power, bond force and time. The effects of these bonding parameters on bondability and ball bond reliability are investigated. We introduce the concept of "reduced bonding parameter" (RBP) to relate these bonding parameters to bondability and reliability. RBP may be regarded as the amount of energy that is put in the ball bond during bonding. Using the appropriate RBP, ball shear force (BSF), ball shear stress (BSS) defined as BSF per unit area, and median ball bond degradation time (t50%) appear to be well-behaved functions of the RBP for a wide range of settings. Optimizing bonding parameter window can be easily performed with this concept.

## 2. Experimental

Bonding with a parameter matrix was done on three-metal layer CMOS bond pads with an opening of 90 μm. BSF and BS measurements were performed prior to molding and packaging. Chips were packaged in a plastic DIL24 for ball bond reliability study.

### *2.1. Bonding parameter matrix*

Table 1 lists the values of the three variables: power, force and time. The combination leads to 60 bonding parameter sets. Bond temperature, bonding machine, ultrasonic frequency, the type of capillary,

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00093-6

material variables including bond pad construction are kept constant in this study.

Table 1
Bonding parameter variables, leading to a total of 60 parameter sets.

| | | | | | |
|---|---|---|---|---|---|
| Power (mW) | 40 | 50 | 60 | 70 | 80 |
| Force (g) | 40 | 50 | 60 | 70 | |
| Time (ms) | 10 | 15 | 20 | | |

### 2.2. *Ball bond bondability*

Ball bond quality is controlled by measuring ball shear force (BSF) and ball size (BS). BSF can offer sufficient sensitivity to reveal the presence of contamination on bond pads, poor metallization adherence or any change in bond pad metallization characteristics. Therefore, BSF is one of the most important criteria for bondability of wire bonding. Ball size is measured through the outer edge of the ball. Ball shear stress (BSS) is then defined as BSF divided by $\pi(BS)^2/4$. This underestimates the real BSS because the real ball bond contact area is smaller than that derived from the measured BS. It has been demonstrated that BSS is an intrinsic property of Au-to-Al metallization, and is a better parameter for bond optimization than BSF [7].

### 2.3. *Ball bond reliability*

Ball bond reliability is determined by ball bond degradation during high temperature storage. Figure 1 depicts a schematic drawing of the test structure for the ball bond degradation measurement. It is a three-point test structure, in which the resistance of the bond pad to the end of a wire is measured. This type of structure can be found in various types of commercial products, which is a very attractive feature of this method, because in this way, one can directly evaluate the ball bond reliability of a real product. For a proper four-point measurement, a double bond needs to be made. As this may influence the bond quality, it is not used in these ball bond integrity studies.

In-situ ball bond degradation monitoring is actually a low current "electromigration" test. The resistance is monitored as a function of time during a very low constant current stress at elevated temperature. The current density is chosen to be low enough to neglect electromigration effects and high enough to allow sensitive resistance monitoring. Only in the end of the experiment when the ball bond is heavily degraded and the contact area becomes much smaller causing the local current density to increase dramatically, then electromigration effect comes into play and a failure is recorded.

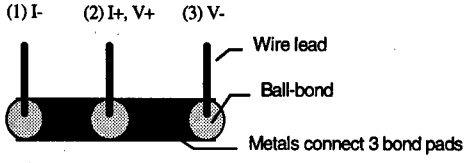



Figure 1 Schematic drawing of the test structure used for the in-situ ball bond degradation measurement. Constant current is applied from (2) to (1) and the voltage change between (2) and (3) is measured. The change in voltage is due to ball bond degradation of (2).

## 3. Results and data analysis

### 3.1 *Ball bond degradation*

A typical ball bond degradation at 250 C° is shown in Figure 2. The sudden increase of resistance reveals the heavily degraded ball bond. Note that a resistance increase of 200% in this case corresponds to a ball bond interface resistance increase of around 2 decades. It is interesting to point out that the ball bond is not completely voided, as indicated by the oscillation of the relative resistance change. The reason is believed to be that when the ball bond is so heavily degraded, the contact area becomes smaller and smaller. At certain moment all current runs through some very small contact areas resulting in connects and disconnects via local electromigration phenomena or even meting.

The failure criterion is 5% increase of resistance. Assuming that devices fail lognormally as a function of time, one can obtain the median time to failure (t50%) and the shape parameter sigma by least-square fit with a line function. Having 60 parameter sets, no large sample size for a parameter set can be used. We use 6 samples for each set. The shape parameter sigma varied largely due to the small sample size. t50% is the most stable parameter for very small populations like these. Therefore, t50% is defined as a measure of the ball bond reliability. Figure 3 shows an example of a lognormal distribution of ball bond degradation at 250 °C.

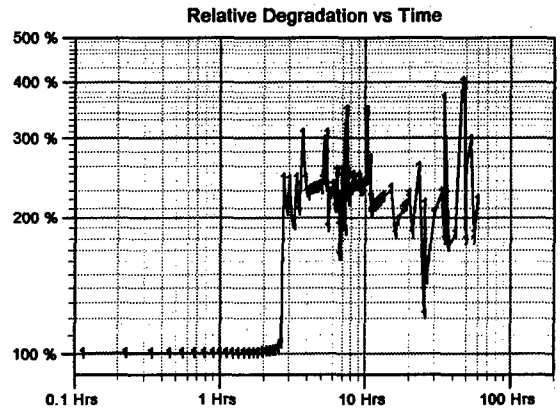



Figure 2 An example of ball bond degradation at 250 C°. The sudden increase of resistance results from heavily degraded ball bond.

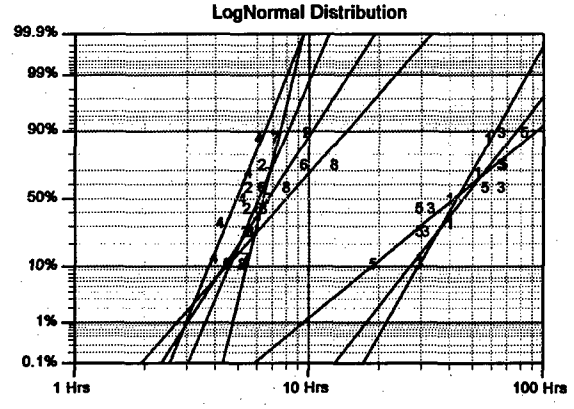



Figure 3 An example of lognormal distribution of ball bond degradation at 250 °C.

### 3.2 Reduced bonding parameter

Multiple regression analysis is often used to fit BSF and BS data. This requires more than 10 fitting parameters to obtain a reasonable fit to the data [5,7]. Furthermore, it is not always clear which bonding parameter has the most significant contribution to the BSF and BS, because it sometimes turns out that an interaction term has the strongest effect from the multiple regression analysis [5]. There is no simple way to plot BSF and BS against bonding parameters, so usually contour plot or 3D chart are used to express the response of BSF or BS as a function of bonding parameters.

We have investigated whether there is a simple analytical function of the three parameters that is related to the BSF. We found a "parameter" that we named reduced bonding parameter and defined it as

$$RBP = Power^{A} \times Force^{B} \times Time^{C},$$

that does exactly what one would like to see: BSF can be transformed into a continuous function of RBP. Through non-linear least square fit with this function to the BSF data, we have obtained the values of A, B and C to be: A= 0.80, B = 0.40 and C = 0.20, with standard deviations of 0.02, 0.02 and 0.03, respectively. As can be seen from Figure 4, BSF is a very well-behaved function of RBP. The RBP can be viewed as the "effective energy" put into the ball bond, that results in Au-Al intermetallics formation and Au ball deformation.

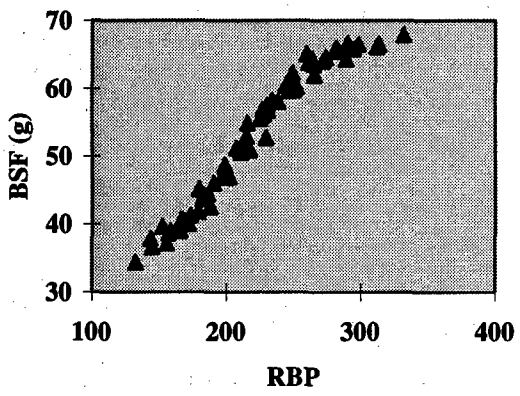



Figure 4 Ball shear force versus reduced bonding parameter. RBP = $Power^{0.8} \times Force^{0.4} \times Time^{0.2}$ (in mW×g×ms).

Although the RBP concept is no physical model but only a phenomenological tool, it provides us a very simple way to manage bonding parameters. It indicates the ranking of bonding parameters on BSF and also allows a calculation of bonding parameter window, as demonstrated in section 3.4. Moreover, instead of more than 10 fitting parameters [5,7], one needs no more than 6 fitting parameters to get the same goodness of fit.

### 3.3 Ball shear stress and median time to failure

With the RBP as x-axis, we can plot BSS ($kg/mm^2$) and t50% (hours), as shown in Figure 5. Since t50% has large variation, in Figure 5 each t50% represents an average t50% of 3 adjacent points. Both BSS and t50% first increase with RBP and reach a maximum at about 255, and then decrease

slightly. It is evident that there is a good correlation between BSS and ball bond degradation The fact that t50% has a large variation is attributed to the relatively small sample size per parameter set.

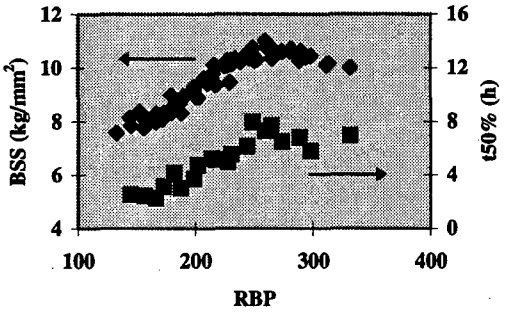



Figure 5 Ball shear stress and median time to failure are plotted against RBP. The correlation between BSS and t50% is readily seen

*3.4 Optimized bonding parameter window*

Because t50% and BSS appear to be correlated, we can use BSS for optimizing bonding parameters. RBP = 255 is the optimized point (Figure 5), which corresponds to a ball size of 86 μm. Taking the constraint that ball size should not be larger than the bond pad opening (RBP = 295), then one may find a window for RBP to be around 255±40. Note that this simple window still gives a lot of freedom for parameter choice. When one remains within the bonding parameter ranges of the experiment, this does not pose a problem.

In practice, bond time is not frequently changed because increasing the bond time decreases throughput. Therefore, ultrasonic power and bond force are the most easily varied parameters in attempts to optimize the ball bond process and reduce bonder variability.

BSS versus RBP data can be fitted with a polynomial function. Using this function we can make a contour plot of BSS as a function of power and force for a fixed bond time, as shown in Figure 6 for a bond time of 15 ms. We note that for the same BSS, force has a larger window than power. This is also revealed by the exponents of force and power in the RBP: namely 0.4 for force and 0.8 for power. In other words, varying power is most effective in changing BSS. It is interesting to note that a window determined by force and power is no longer a rectangular with the RBP concept.

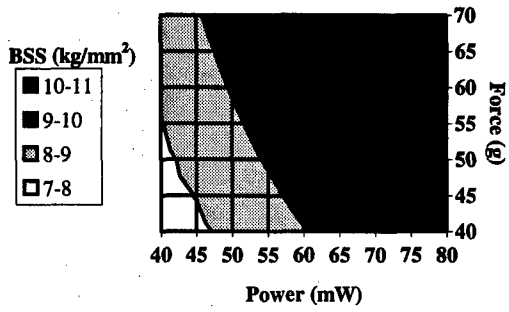



Figure 6 Contour plot of ball shear stress as a function of ultrasonic power and bond force. Bond time is 15 ms.

## 4. Conclusions

We have studied the effects of bonding parameters on bondability and ball bond reliability. Bondability is better described using ball shear stress and ball bond reliability can be determined by in-situ ball bond degradation measurements. By introducing the concept of reduced bonding parameter, we are able to relate the bonding parameters directly to bondability and reliability. Using the appropriate RBP, ball shear force (BSF), ball shear stress (BSS), and ball bond degradation time (t50%) appear to be well-behaved functions of the RBP for a wide range of settings. This provides us with an analytical tool for obtaining bonding parameter window.

## Acknowledgements

We thank Gidi Fleuren, Jan van Kempen for their contributions to this work. K. C. Kuo is acknowledged for support and discussions.

## References

[1] T. Uno, K. Tatsumi, and Y. Ohno, "Void formation and reliability in gold-aluminum bonding", Adv. in Electronic Packaging, ASME, 1992, p 771.

[2] W. Gerling, "Electrical and physical characterization of gold-ball bonds on aluminum layers", Electronic Components Conf., 1984, p 13.

[3] D. T. Rooney, J. B. Dixon, K. G. Schurr and M. R. Northrup, "Evaluation of wire bond quality by SEM analysis of ball-shape and visual inspection of

intermetallic formation", ISHM'96 Proceedings, 1996 p 432.
[4] T. D. Hund and P. V. Plunkett, "Improving thermosonic gold ball bond reliability", IEEE Trans. Comp., Hybrids, and Manuf. Technol., CHMT-8 (1986) 446.
[5] P. M. O'Neill, D. H. Eaton, T. Phua, F. Loh, and B. Chin, "Wirebonding to multilevel metal", Solid State Technology, 40 (1997) 87.
[6] Y. S. Chen and H. Fatemi, "Au wire bonding evaluation by fractional fractorial designed experiment", The International Journal for Hybrid Microelectronics, 10 (1987) 1.
[7] W. K. Shu, "Fine Pitch Wire Bonding Development Using Statistical Design of Experiment", IEEE 1995 Proceedings. 45th Electronic Components and Technology Conference, 1995, p 91.
[8] M. Sheaffer and L. Levine, "How to Optimize and Control the Wire Bonding Process: Part I", Solid State Technology, , Nov. 1990, p 119.
[9] M. Shell-De Guzman and M. Mahaney, "The bond shear test: An application for the reduction of common causes of gold ball bond process variation", IEEE/IRPS, 1992, p 251.

PERGAMON

Microelectronics Reliability 38 (1998) 1293–1296

MICROELECTRONICS RELIABILITY

# Measurement of the thermomechanical behaviour of the solder-lead interface in solder joints by laser probing : a new method for measuring the bond quality.

S. Dilhaire, A. Cornet*, E. Schaub, C. Rauzan and W. Claeys

Laboratoire de Caractérisation de Composants Electroniques, CPMOH and IXL
Université Bordeaux 1, 351 cours de la Libération 33405 Talence Cedex, France
Fax : 33(0) 556 846 970, E-Mail : wclaeys@frbdx11.cribx1.u-bordeaux.fr
* Laboratoire FYAM, Université de Louvain, 1348 Louvain-La-Neuve, Belgium.

**Abstract**

We present an original method of laser probing to observe thermomechanical stresses at interfaces of solder joints. The stress is generated by Peltier heat exchange and observed through thermal expansion in the picometer scale. A simple model allows to extract a parameter measuring the interface bond quality.

© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

The quality and reliability of solder joints is a critical issue in microelectronics. Many methods of inspection and accelerated ageing tests have been proposed [1,2]. For instance, we have used a homodyne stabilised interferometer [3], specially designed for applications in microelectronics, to study the ageing of solder joints [4,5,6]. We followed its quality evolution during ageing tests through the analysis of its electro-thermo-mechanical behaviour.

Concerning failure mechanisms occuring in solder joints, it has been shown that failure is mainly due to cracks at interfaces [1]. The cracks originate from thermomechanical stresses generated by the difference between the coefficients of thermal expansion of joint materials. They lead to an irreversible degradation of the bonding between layers. That's why the information about the initial quality of the interfaces is of crucial importance. But no quantitative study of this kind was reported.

In this paper we present an original approach to analyse the thermomechanical behaviour of the solder-lead interface in solder joints by laser probing. Local (micrometric lateral resolution) thermomechanical strain produced by the Peltier effect when the joint is current crossed at solder-lead interface is measured by laser probing upon a solder-joint microsection.

The method is promising for the non contact analysis of the action of thermomechanical stress inside metallic bonds. The method can serve as an initial quality determination after process and can also serve in accelerated ageing tests to follow the evolution of the bond quality.

## 2. Principle of the method

Many problems related to quality and reliability of solder joints are due to the thermal expansion mismatch between the different metals involved. This mismatch produces residual stress during the cooling period of the fabrication process. Thermal cycles and vibrations generate voids and cracks in the region of residual stress, this is at the interface between different metals. It is for these reasons that we propose a method that analyses the quality of the bond at the metal interfaces in solder joint.

The principle of the method is to measure the thermal expansion resulting from a local thermal perturbation at the lead-solder interface. The thermal expansion mismatch is maximum in this region and we explore over a distance where the effect of this mismatch can be observed.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00080-8

We measure thermal expansion due to Peltier heat exchange produced by current driven through the interface. Figure 1 illustrates the method. A parallelepiped of solder material is sketched showing the interface of two metals, A is the lead and B the solder. A current is driven from A to B in the y-direction through the interface. Peltier heat is produced at the interface (xz plane) and diffuses within A and B in the y direction producing a temperature distribution in both metals. We measure by laser probing on top of the sample the thermal expansion in the x-direction resulting from this temperature distribution at different locations along y. This thermal expansion $\Delta x$ can be expressed, due to symmetry reasons, as the product of the temperature variation T(y), the metal thickness $X_0$ and an effective local thermal expansion coefficient $\alpha_{eff}(y)$:

$$\Delta x = X_0 . \alpha_{eff}(y) . T(y). \qquad (1)$$

$\alpha_{eff}(y)$ is a phenomenological quantity which reflects the influence of the mismatch of thermal expansion between A and B. The region over which the mismatch is observed shows where the thermal stress acts. $\alpha_{eff}(y)$ will vary from $\alpha_A$ (y<<0) to $\alpha_B$ (y>>0), where $\alpha_A$ and $\alpha_B$ are the free thermal expansion coefficients of respectively metal A and B. The extension of this region, for a given thermal perturbation, depends upon the quality of the bond. One readily understands that a loose bond will produce a step function for $\alpha_{eff}(y)$, jumping from $\alpha_A$ to $\alpha_B$ at x=0.

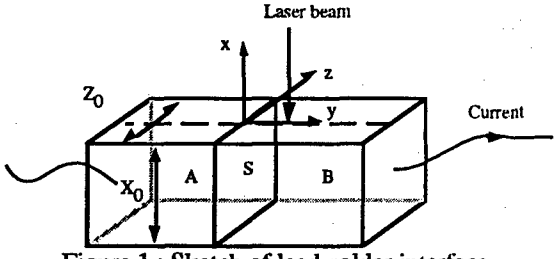



Figure 1 : Sketch of lead-solder interface.

## 3. Thermomechanical study of a solder joint

Measurements are performed upon the microsection of a solder joint. A sine wave current is fed to the device (which is kept functional) as shown in figure 2. According to the Peltier effect, the solder - lead interface becomes a heat exchanger and generates a thermomechanical field in the surrounding media. Together with Peltier heat, Joule heat is produced inside the metals where current flows. To discriminate Joule temperature changes from the

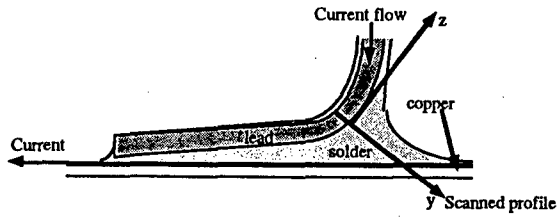



Figure 2 : Microsection of a solder joint.

Peltier ones, we measure the first harmonic thermal expansion, as the Joule heating temperature variations are produced at the second harmonic [3,4,5,6]. The induced normal surface displacement is detected by a high resolution interferometric laser probe [3].

The absolute value of the surface displacement amplitude is shown in figure 3 for different positions along the y-axis perpendicular to the interface plane (see figure 2). The magnitude of the displacement is of the order of $10^{-12}$ meter. The y = 0 location is at the lead-solder interface.

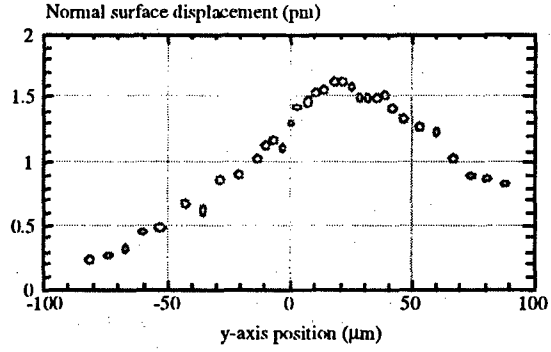



Figure 3 : Measured normal surface displacement showing the maximum not to coincide with the heat source (y=0)

The most important feature is that the maximum of the deformation does not coincide with the position of the heat source (highest temperature) as should be the case for a homogeneous medium. The maximum temperature variation is due to Peltier heat exchange and is located at the metal interface (y=0). In our case, the shape of the deformation profile depends not only upon the temperature distribution but also upon the elasticity properties of the joint. This shift is due to the mismatch of thermal expansion coefficient of the two materials. The solder expansion coefficient is five time larger than the lead one. A stress field is generated in the contact region. As a result we have to deal with an effective expansion coefficient taking all the thermal stresses present in the media into account. This explains why the maximum deformation is observed in the solder.

## 4. Simple model for strain distribution

We propose a very simple analytical model in order to give a qualitative explanation of the measured profile. Two pieces of different metals A and B having the same thickness $X_0$ are bond together over a section $S=X_0.Z_0$ (see figure 1). The laser probe measures the surface x displacement along the Oy direction.

By applying a sine wave current to the structure we have a surface periodic heat exchanger at the interface S. For a 10 kHz excitation frequency the thermal wave diffusion length is of the order of a few ten microns. This value is much smaller than the dimensions of the device under test namely $X_0$ and $Z_0$ such that the source (plane S) can be considered as an infinite plane. For simplicity, we suppose the temperature to be uniform in the Ox and Oz directions, only variable y remains. Under these conditions the temperature distribution is well known [7] and can be expressed:

$$T(x,y,z)=T(y)=T_0 e^{-\frac{y}{L_{TH}}} e^{i\left(\omega t-\frac{y}{L_{TH}}\right)} \text{ for } y>0 \quad (2)$$

$$T(x,y,z)=T(y)=T_0 e^{+\frac{y}{L_{TH}}} e^{i\left(\omega t+\frac{y}{L_{TH}}\right)} \text{ for } y<0 \quad (3)$$

$L_{TH}$ is the thermal diffusion length. Figure 4 shows the temperature distribution in the y direction.

The normal surface displacement can then be expressed

$$\Delta x(y)=\int_{-X_0}^{0}\alpha_{eff}(y).T(y)dx$$

$$=\begin{cases} X_0\alpha_{eff}(y)T_0 e^{+\frac{y}{L_{TH}}} e^{i\left(\omega t+\frac{y}{L_{TH}}\right)} & \text{for } y<0 \\ X_0\alpha_{eff}(y)T_0 e^{-\frac{y}{L_{TH}}} e^{i\left(\omega t-\frac{y}{L_{TH}}\right)} & \text{for } y>0 \end{cases} \quad (4)$$

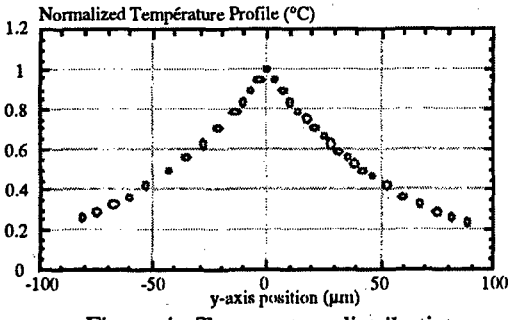



Figure 4 : Temperature distribution.

$\alpha_{eff}(y)$ is the phenomenological thermal expansion coefficient taking all the thermal stresses present in the materials into account.

To obtain the $\alpha_{eff}(y)$ profile we divide the measured profile of surface displacement by the temperature one. For simplicity we assume the temperature profile to be described by equation 2 and 3 presented in figure 4. The temperature is normalised with respect to the maximum value. The result of the division is shown in figure 5, where $\alpha_L$ is taken equal to 1.

In the contact region, instead of a step change, we observe a continuous change in effective expansion coefficient. Far away from the interface, the $\alpha_{eff}(y)$ tends to constant values:

$$\alpha_{eff}(+\infty)\rightarrow\alpha_S \quad ; \quad \alpha_{eff}(-\infty)\rightarrow\alpha_L \quad (5)$$

As it can be expected they should be the free expansion coefficients of metal A (solder) and B (lead). The mutual mechanical influence between A and B is localised in the contact region.

The effective expansion coefficient evolves from $\alpha_S$ to $\alpha_L$ in the interface region.

To verify this we extract from the results the ratio between the two effective expansion coefficients at $y=\pm\infty$, which is $\frac{\alpha_S}{\alpha_L}\approx 4$. This result is in a good agreement with the one expected from literature [8].

From the data of figure 5 we derive a very important quantitative phenomenological constant of the bonding which is the extension of the stress region ‹ρ› (figure 5), we define it as the length over which 90 % of effective thermal expansion coefficient variation occurs. This means the distance between the points both in the lead and the solder where the thermal expansion has varied by 5% of its total change. It is of the order of 100 µm for the case of the solder joint we have studied. It is obvious that

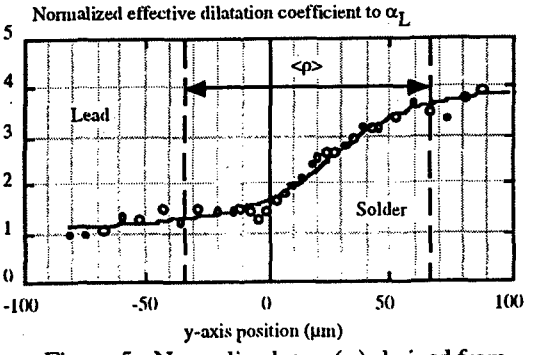



Figure 5 : Normalized $\alpha_{eff}(y)$ derived from calculation.

this range will go to zero when the bonding gets loose as it is the case where both A and B can expand freely. The extension ‹ρ› becomes therefore a good evaluation of the bond rigidity for a given excitation frequency.

## 5. Conclusion

We have presented for the first time to our knowledge, a quantitative analysis of interface stresses in solder joints. We have proposed a simple model to explain qualitatively the observed result. This opens a new way to study the influence of stresses upon the quality of solder joints under operating conditions. The extension ‹ρ› of the stress region centred at the lead-solder interface could be used to evaluate the strength and the rigidity of the interface bond.
For durther studies it should be interesting to study the relationship of ‹ρ› with respect to the joint lifetime.

## 6. References

[1] A. Hijazi, Y. Danto, Y. Ousten and B. Allamargot, *Proceedings of ESREF 92*, Schwäbisch-Gmünd, Germany, p. 327, (1992).
[2] R. Vanzetti and J. S. Hwang, *NEPCON WEST* (1989).
[3] W. Claeys, S. Dilhaire and V. Quintard, *Microelectronic Engineering*, 24, p. 411-420, (1994).
[4] W. Claeys, S. Dilhaire, V. Quintard, D.Lewis and Y. Danto, *Quality and Realiability Engineering International*, Vol 10, 237-242, (1994).
[5] W. Claeys, V. Quintard, S. Dilhaire, D.Lewis, T.Phan and Y. Danto, *Quality and Realiability Engineering International*, Vol 10, 289-295, (1994).
[6] V. Quintard, B. Parmentier, T. Phan, D. Lewis, S. Dilhaire and W. Claeys, *Quality and Reliability Engineering International*, Vol 12, 447-451, (1996).
[7] Carslaw W. and Jaeger H. S, *Conduction of heat in solids*, Clarendon Press - Oxford 3d edition, (1994).
[8] Handbook of Chemistry and Physics, 52nd edition, The Chemical Rubber Co, (1971-1972).

PERGAMON

Microelectronics Reliability 38 (1998) 1297–1300

MICROELECTRONICS RELIABILITY

# Experimental design and evaluation of interconnection materials for improvement of joint reliability at power transistors

P. Jansson

*Ericsson Radio Systems AB, S-800 06 Gävle, Sweden*

**Abstract**

Experimental design methodology has been used in order to find parameters that affect the solder joint reliability of radio frequency transistors in radio base stations for mobile telecommunication. The design of component leads and different interconnection materials have been studied. The solder composition Au80Sn20 showed to increase the life time to a large extent. Thickness of the solder joints has a great influence on reliability and if that parameter is well controlled the solder Sn25Pb62Ag03Bi10 can be a good alternative. The effect of gold stripping was also investigated. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

In the transceiver modules in radio base stations for mobile telecommunication, radio frequency transistors are important devices. These transistors transmit power of typically 10, 30 or 50 W and are submitted to extreme temperature cycling since the power amplifier is frequently switched on and off.

Failure analysis has concluded that solder joint fatigue causes interconnection ruptures between transistor lead and board. Therefor, the present work started with the goal to identify parameters that affect the joint reliability and to give recommendations for the assembly process (normally soldering).

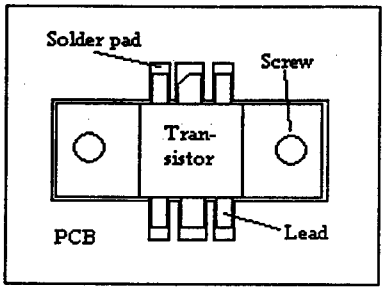



**Fig 1.** Power transistor on PCB

Figure 1 shows a schematic view of a typical transistor mounted on a printed circuit board (PCB). The six leads are soldered to the PCB and the flange is screwed to an aluminum casting.

## 2. Procedure

The strategy for the work was to use a Design Of Experiment (DOE) methodology, after evaluation make some changes to the design or process and finally verify the result. The first experiment was passive thermal cycling but for the verification accelerated radio frequency power cycling was used.

### *2.1. Description of factors and passive cycling*

After brain storming, a test matrix consisting of 16 experiments and 9 factors was created.

*Interconnection material* between transistor lead and PCB. Two solders, Sn62Pb36Ag02 and Sn25Pb62Ag03Bi10, and two isotropic conductive adhesives were tested.

*Interconnection height.* The distance between lead and PCB was varied, i.e. the solder joint thickness. Difference between "low" and "high" was 0.5 mm.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00090-0

*Lead form.* The leads were either bent or straight, see figure 2.



**Fig2.** Bent lead

*Slits or not.* Two slits were made in the lead, see figure 3.

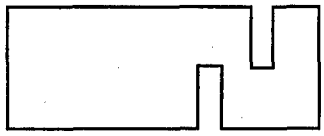


**Fig 3.** Two slits in a lead

*Lead length.* The two different lengths were approximately 2.5 and 4.0 mm.

*Maximum temperature.* Peak temperature in the temperature cycling was 110°C or 130°C but with the same ΔT.

*Lead material* was either copper or Alloy 42.

*Number of screws* holding the PCB. The choice was with or without extra screws near the transistor.

*Transistor flange mount.* Some of the transistors were fixed to the casting with adhesive instead of screws.

The temperature cycling test conditions are described in Table 1.

Table 1. Test condition for passive thermal cycling.

| Parameter | Data |
|---|---|
| ΔT | 105°C |
| Cycle time | 2 h |
| Dwell time | 15 min |

Visual inspection of the solder joints and resistance measurements were performed after 0, 60, 120,..., 540 cycles. Failure criteria was visual recrystallisation at the lead tip, cracks or 10% increase in resistance.

### *2.2 Radio frequency power cycling*

Since a power cycling test is more realistic compared to the real service behaviour of the product, the most interesting combinations of transistor design and process parameters were studied in such test. The present design (short, straight leads with 3-4 μm gold on the surface and soldered with Sn62Pb36Ag02) was included as reference specimen. Table 2 describes the power cycling test set up.

Each power amplifier unit contains two transistors and four such units were tested for each experiment in Table 2. The arrangement is described schematically in Figure 4.

No external heating or cooling equipment was used to create the temperature swing. In order to accelerate the test the in-built cooling fan was turned off when the power transistor was on and the fan was turned on at power off. The cooling flanges on the casting had to be milled off. This resulted in a temperature change (ΔT) of 80°C. The cycle time was 30 minutes, i.e. 15 minutes at power on and 15 minutes at power off.

Power amplifier malfunction was set as failure criteria and visual inspection of interconnection joints was used for verification of the failures.

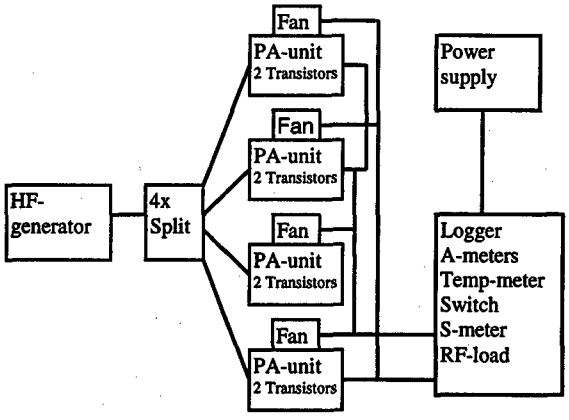



**Fig 4.** Schematic drawing of the test system.

Table 2. Test cases in the power cycling test

| No | Solder/Adhesive | Lead length | Lead shape | Lead finish | Joint thickness | PCB finish |
|---|---|---|---|---|---|---|
| 1 | Au80Sn20 | Short | Straight | Au | Standard | SnPb |
| 2 | Sn25Pb62Ag03Bi10 | Long | Bent | Au | >200 μm | SnPb |
| 3 | Adhesive | Long | Straight with hole | Au | Standard | Ni/Au |
| 4 | Sn25Pb62Ag03Bi10 | Long | Bent | Au | Standard | SnPb |
| 5 | Sn25Pb62Ag03Bi10 | Short | Straight | Au stripped | Standard | SnPb |
| 6-7 | Sn62Pb36Ag02 | Short | Straight | Au | Standard | SnPb |
| 8 | Sn62Pb36Ag02 | Long | bent | Au stripped | Standard | SnPb |
| 9 | Sn25Pb62Ag03Bi10 | Short | Straight | Au | Standard | SnPb |
| 10 | SnPb + Ni-spheres | Long | Bent | Au | Standard | SnPb |
| 11 | Adhesive | Long | Straight with hole | Au | Standard | SnPb |

## 3. Test results

### *3.1 Passive thermal cycling*

This experiment and the results are reported earlier in [1]. Important conclusions were:

- By using isotropic conducting adhesive and longer leads, the life should increase by a factor 4.5.
- Using the adhesive and present (short) leads increase the life by a factor 3.5.
- The solder Sn25Pb62Ag03Bi10 combined with long and bent leads increase the life by a factor 2.3.
- The solder Sn25Pb62Ag03Bi10 and keeping the present leads (short and straight) increase the life by a factor 1.5.
- Alloy-42 or Cu as lead material does not affect the solder joint life.

These statements were important input for design of the power cycling test.

### *3.2 Result of power cycling*

The test results from power cycling can be seen in Figure 5. To make the graph, the method of censored data was used. This was necessary as we took away one working transistor each time the other transistor on the amplifier unit failed.

The cycling of the units according to test case 1 was interrupted after 6470 cycles. No failures were detected during the test period.

### *3.3 Discussion of results*

The results for the two different solders are contradictory. Sample 9 (Sn25Pb62Ag03Bi10) showed shorter life than samples 6 and 7 with Sn62Pb36Ag02. Cross sections from these three groups showed that the differences in solder joint thickness are big. The reason is probably, that manual soldering process was used with difficulty in the control of solder height. The thickness interval was approximately 10-100 μm. There was also a tendency of increased formation and propagation of cracks in samples with thin joints.

It seams that the joint thickness is a parameter that strongly affect the fatigue behaviour. Group 2 and 4 support that theory since these samples are identical except the joint thickness. The stand-off was not allowed to be below 200 μm in group 2, which showed better result than group 4.

The boards with adhesive instead of solder showed extremely short life compared to the other

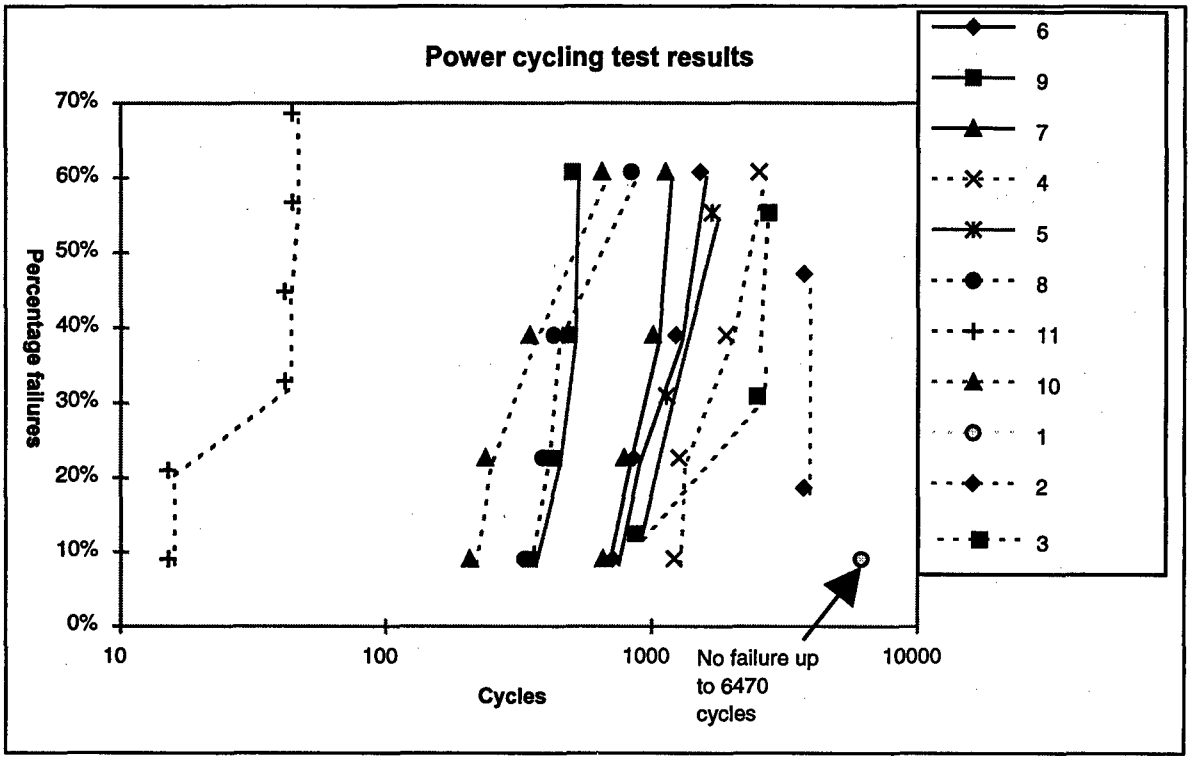



**Fig 5.** Results from accelerated power cycling

groups when SnPb was used as PCB finish. The result was among the best when boards with Ni/Au were used. These boards were cleaned with isopropanol before the adhesive was applied. The spread in the result is in spite of that larger for adhesive compared to solder.

Discussions whether the gold on transistor leads shall be removed or not prior to soldering is frequent. This study does not support the theory that gold strip is necessary.

## 4. Conclusions

Using Au80Sn20 solder instead of ordinary near eutectic SnPb alloys would probably improve the reliability by a factor of at least 6. But, since the melting temperature is 280°C a suitable soldering process has to be developed.

If the solder joint thickness is under control, changing to Sn25Pb62Ag03Bi10 could be a good solution too. The influence of the thickness seems so big that the effect of other factors are difficult to detect.

Removing the gold from transistor leads prior to soldering is not necessary and can cause other problems, such as electrostatic discharge damages for example. Solder joints shall have a thickness of some order to secure that brittle intermetallic compounds do not increase the risk for crack propagation.

## References

[1] Olsson C. Experimental design for improvement of stressed solder joint reliability. ESREF –97 in Bordeaux.

PERGAMON

Microelectronics Reliability 38 (1998) 1301–1305

MICROELECTRONICS RELIABILITY

# Crack Mechanism in Wire Bonding Joints

S. Ramminger *, P. Türkes *, G. Wachutka **

* *Siemens AG, Corporate Technology Department, Munich*

** *Institute for Physics of Electrotechnology, Technical University of Munich*

**Abstract**

High voltage and high current power modules are key components for traction applications. While the modules are exposed to harsh stress conditions all over their lifetime, high reliability is of decisive importance in this field of application. In power electronic packages wire bonding is used for the electrical interconnection from the chips to the output pins. Wire bond lift-off and solder fatigue are limiting the reliability. In this work we investigate the initiation and growth of cracks in the wire bonds using finite-element analysis.
© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Insulated Gate Bipolar Transistor (IGBT) packaged in high power modules with blocking voltages up to 3.3 kV and current ratings up to 1200 A are favored devices for traction and many other industrial applications. Up to now, for the traction converters in urban systems and in high speed trains diodes, thyristors and mainly GTO-thyristors (Gate turn off) with the press-pack-technology were used. By now, high power IGBT-modules manufactured in packaging technology, using internal soldered electrical contact and wire bonds instead of pressed contacts, appear in high power traction applications.

The main advantages offered by these modules are the low on-state resistance, the high input impedance, and high switching capability which make them superior to the traditional high-voltage and high-power switching components such as thyristors.

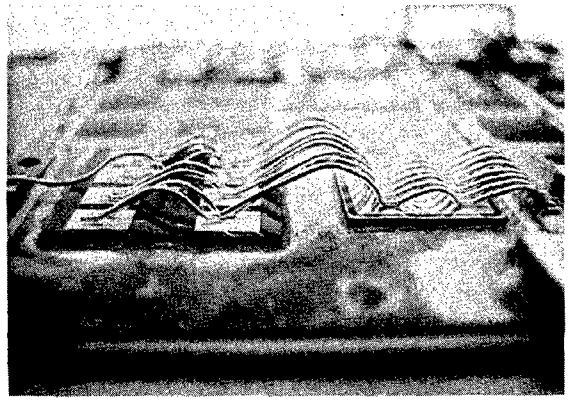


Figure 1. Wire connections between the IGBT, the freewheeling diode and the output pins in a power module

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00141-3

A further improvement is the module integrated free-wheeling diode. The technology of module packaging solves mechanical mounting and isolation aspects. The IGBT combines the advantages of a power bipolar transistor with those of a power MOS transistor. At least, the cost of module packaging is lower than the press pack packaging because of a very large production for the industrial applications.

In spite of their many advantages, the reliability of IGBT modules still involves serious concerns. To know more about the life expectancy of IGBT modules a lot of researcher groups have done power cycle tests. In such a test shorter cycle periods were applied to the module than in real applications because of the time-saving aspect for the investigation of the failures. Today it is a well-known fact that, among others, wire bonding and die attachment are the most limiting factors for the reliability of power modules.

## 2. Wire bonding

It is typical of the packaging of power IGBT modules that several heterogeneous materials are joined together by different methods. The active silicon devices are soldered to a DCB ceramic (Direct Copper Bonding) which, in turn, is soldered to the copper base plate. A main problem in packaging is the thermomechanical stress caused by temperature cycling during service life of the device due to the different thermal expansion coefficients (CTE) of the layers.

The design of a high power module requires a parallel connection of up to 24 chips. The chips are soldered to one or more DCB substrates (Direct copper bonding) which facilitate the electrical isolation from the environment. This substrates are soldered to a copper baseplate which represents the heat sink during power application. High purity aluminum wires (99.999%) with diameters up to 500 μm connect the silicon device with the output pins of power electronic packages. It is necessary to use several wires to handle rated currents up to 75 A per chip. In the on-state mode, the maximum electrical current is almost 10 A per wire. Figure 1 shows the wire bonds in a power module. Each wire is bonded twice on thin aluminum pads (some μm thick) on the dies.

The deformation process during wire bonding is rather complex. The bonding tool is pressed on the wire and produces an ultrasonic vibration in the longitudinal direction of the wires. The wire breaks the aluminum oxide film and is welded to the pad. Understanding the bonding mechanism is essential for the determination of the external parameters. The most important parameters are the ultrasonic energy (7.5-8.5 W) and the frequency (6 kHZ), the bonding force (600-750 cN) and time (150-200 ms) [1]. The quality of a bond can be controlled in a pulltest.

## 3. Failure mechanism

High reliability is the most important requirement. The reliability of any electronic device module depends on a great part on its construction. One fast test method to investigate the reliability of the modules is a power cycling test [2]. The thermal mismatch between the wire bond and the chip generates a significant thermomechanical stress in the bonding zone during temperature cycles.

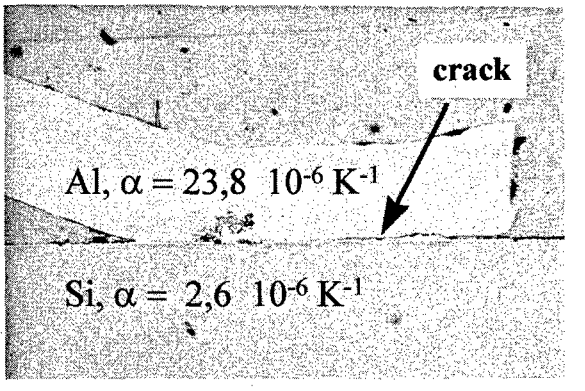



Figure 2. Cross section of an aluminum wedge bond with a crack in the bonding zone

The main failure which is observed in power modules is solder joint fatigue and wire bond lift-off [3]. Wire bond connections are susceptible to fatigue as a result of thermomechanical damage mechanism during power operation. This paper presents new results about the bond wire lift-off.

Figure 2 shows a typical failure in an aluminum wedge bond after thermocycling. A crack is observed between the wire and the metallization. The highest thermomechanical stress occurs at the wire terminations, where the wire is unable to flex. Consequently, the typical failure sites are the heel and the tail of a wedge bond. The wire is unable to accomodate the locally caused strain at this place and a crack is initiated. During the bonding process, the wire material is strongly deformed and small grains are built in the interface region during the following system soldering process. The grain size in the wire center is much larger than in the welded zone. There are mainly two areas of different grain sizes. Under thermomechanical stress, cracks grow from the heel and the tail along the boundary of fine and coarse aluminum grains into the welded interface [4].

Figure 3 shows a SEM photograph of the lift-off pattern of a bond area. Residuals of wire materials are found on the pattern perimeter at the fracture surface on top of the metallization. No wire material is found in the middle of the bonding zone. This crack behavior gives evidence that the center of the bond area remains unwelded and a strong bond is only formed at the perimeter.

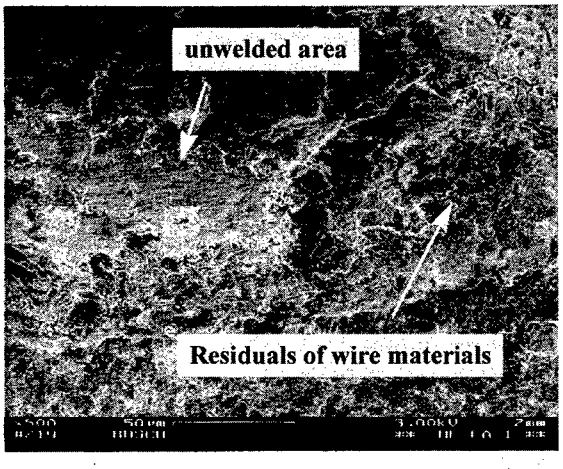



Figure 3. SEM photograph of the bonding area after bond wire lift-off

The unwelded area in the center is assumed to be a result of the wire deformation process. When a crack reaches the center, the wire lifts and the mechanical as well as the electrical contact is interrupted. Unfortunately, this is an self-accelerating process because the remaining wires have to conduct more current causing higher heat dissipation in the wire and in the welded area which again accelerates lifting of further wires.

## 4. Finite element simulation

Crack initiation due to fatigue is an important aspect of materials performance. The terminations of joints are always predestinated for high stress concentration under stress conditons. The initiation and growth of cracks are simulated on the basis of a physical model using the finite-element method (FEM).

The simplified model only consists of the wire (300 μm diameter), the bonded area with its fine grains, the metallization (3 μm thickness), and the device (220 μm). Figure 4 shows geometry and material behaviors of the different layers. The metals have linear-elastic, ideal-plastic behavior with different yield strengths. Because of the small grains, the welded area suffers from higher yield stress as the wire [5], which is estimated to 20 MPa.

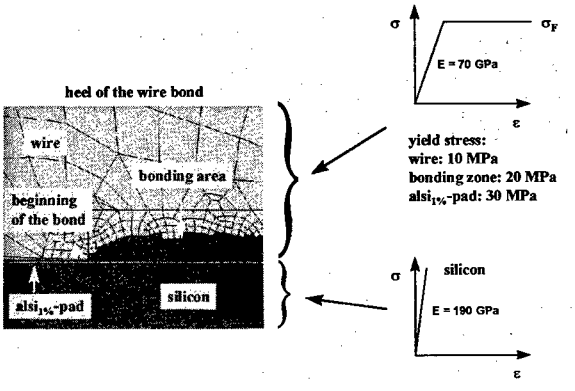



Figure 4. Material behavior and geometry design of the interface area

The behavior of the metallization is unknown but the yield resistance should be higher than that from the wire because of the doping of the bonding pad. The silicon, of course, has only linear elastic behavior with Young's modulus of 190 GPa.

In the initial state, the whole FEM-model is free of mechanical tension at T=100° C. Then, during simulation, the temperature is cooled down with a step of 75 K instantaneously. This causes

stress in the structure induced by the different thermal expansion coefficients. It is evident that when the thermal strain exceeds the elastic limit in the wire, the stress is relaxed by plastic deformation. A failure criterion should be used for the stress results. The well-known stress-intensity factor concept is not applicable to a non-linear simulation with large expansion. Moreover, this concept is not defined at interfaces where the materials are changing their properties.

The J-integral method is used as a criterion for non-linear fracture mechanics. In this special case, however, the method is not applicable because of the non-parallel crack surfaces and the material changes at the crack tip. Furthermore, the J-integral can only handle crack initiation and not crack growth.

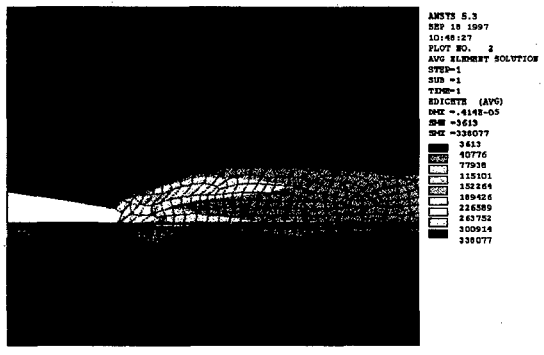



Figure 5. Calculated energy density

Figure 5 shows the calculated energy density after the cooling step. Following the empirical assumption that a high amount of dissipated energy at a small location causes damage in the microstructure, the energy density provides a sufficient failure criterion. The free energy is required to built new surfaces.

So it is assumed that finite elements at locations where the locally inelastic dissipated energy exceeds a certain limit, indicate failure initialization. Hence, these elements are removed in the next calculation steps. Figure 6 shows the calculated crack path after the second, the forth, and the seventh simulation step.

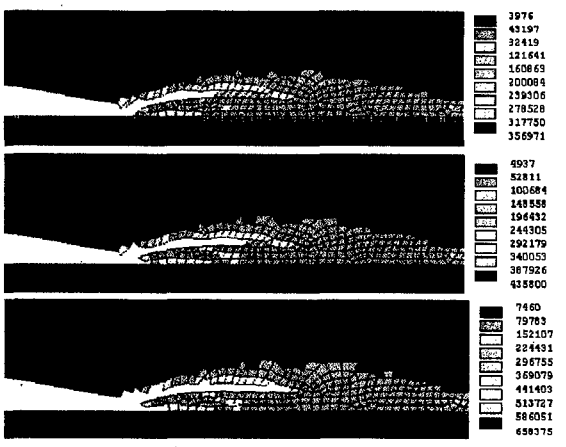



Figure 6. Calculated crack growth at the heel side of the bond with the energy density of 0.18 J/cm$^3$ as a failure criterion

In this passive stress simulation the crack growth stops after a few cycles due to the fact that the bonded joint gets smaller and smaller, that therefore the stress decreases, and that consequently the energy density does not satisfy the failure criterion any more. In an active power test the crack growth will not stop because of the increasing current density and the increasing temperature. Nevertheless, it is possible to calculate the crack initiation and the observed crack path by this numerical experiment. Further simulations with more accurate data of materials are still in progress.

## 5. Conclusion

In power electronic packages, wire bonding is used for the electrical connection between the chips and the output pins. High reliability has uppermost priority. In this work we investigated the cause of power module bond degradation and subsequent failure under power cycling conditons. Unfortunately, this proves to be the most crucial factor for the longterm reliability of power modules. In numerical tests, the observed cracks can be reproduced with a physically plausible, but heuristic model using finite-element calculations. It makes use of a local failure estimator which can describe the initiation and growth of a crack in a qualitatively correct way. Parameter extraction for

the quantitative adjustment of the model parameters is currently in progress.

## References:

[1] J. Hauke, L. Berchtold: "Oberflächen für das Bonden mit Al-Dickdraht", PRONIC 5, 1993

[2] P. Jacob, M. Held, P. Sacco, W. Wu; "Reliability Testing And Analysis Of IGBT Power Semiconductor Modules", Coll. "IGBT Propulsion Drives" pp. 4/1-5, 1995

[3] M.-H. Poech, K. J. Dittmer, D. Gäbisch; "Investigations on the damage mechanism of aluminum wire bonds used for high power applications"; Proc. EUPAC 96, pp. 128-131, 1996

[4] M. Ciappa, P. Malberti; "Plastic-Strain Of Aluminum Interconnects During Pulsed Operation Of IGBT Multichip Modules", Qual. Rel. Eng. Int., Vol. 12, 297-303, 1996

[5] M. Held, P. Jacob, G. Nicoletti, P. Scacco, M.-H. Poech; "Fast Power Cycling Test for IGBT Modules in Traction Application", Proc. PEDS 97, pp. 425-430, 1997

PERGAMON

Microelectronics Reliability 38 (1998) 1307–1312

MICROELECTRONICS RELIABILITY

# Materials interfaces in flip chip interconnects for optical components; performance and degradation mechanisms

Robert H. Esser[a], A. Dimoulas[a], N. Strifas[a], A. Christou[a], N. Papanicolau[b]

[a] *Department of Materials and Nuclear Engineering And Center for Optoelectronics Devices, Interconnects and Packaging, University of Maryland, College Park 20742*

[b] *Naval Research Laboratory, Washington, DC 20375*

**Abstract**

Flip chip solder joint reliability is dependant on final microstructure. The interface between the solder and the underlying metallization is of primary concern since the majority of joint failures occur at or near the interface. We report the results of our investigation on the structure-property relationships of the as deposited metallization and electroplated Pb/Sn solder, as they affect failure mechanisms. Samples investigated were prepared by electron beam evaporation of a Cr/Cu/Au metallization. Electroplated solders were prepared using a high tin solder bath and current density of 3.2mA/cm$^2$. It is shown that interdiffusion and intermetallic formation between gold and copper occurs during evaporation deposition. It is also shown that a significant amount of interdiffusion and intermetallic formation between the tin in the solder and the seed layer metals (Cu and Au) occurs during the electroplating operation. It is believed that this is due largely to resistive heating of the samples. The degree of interdiffusion in the as deposited state determines the ultimate reliability of the flip chip solder bumps in an optical configuration.

© 1998 Elsevier Science Ltd. All rights reserved.

## 1.0 Introduction

Flip chip solder die attach schemes have many applications in the microelectronics industry. The most common application to date is high density interconnection between the chip and the substrate [1,2,3]. A new application for solder ball bonding is in the area of vertically integrated free space optical interconnects [4].

Current microelectronic architectures are two dimensional. A vertically integrated structure allows expansion and utilization of frequently unused substrate area and expands interconnect connectivity to a volumetric one. The predominant concern with vertically integrated free space optical interconnects is planar alignment [5]. Flip chip solder bonding technology can be applied to ensure planar alignment. The passive alignment achieved with properly designed flip chip structures is an adequate solution of the alignment problem [4].

In order for free space optical interconnects to meet specifications and requirements successfully, they must be aligned to extremely close optical tolerances. Flip chip solder interconnect technology has a self centering effect [4,5] that allows the alignment to be done passively. The passive alignment is almost entirely due to the surface tension of the solder [5]. As long as the solder can wet adequately over the majority of the interconnect sites, then the surface tension will provide a horizontal passive alignment. The alignment may be degraded by surface tension

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00131-0

variations due to the presence of intermetallic compounds.

In order to realize the many benefits of optical interconnects, they must be reliable enough to rival current electrical interconnection architectures. Of great concern in flip chip solder bonding reliability is thermal fatigue [6-10][1]. The thermal expansion mismatch between the substrate and the other chips in the module will result in mechanical strain on the solder bumps. As the module experiences a normal heating and cooling cycle during use, the solder joints experience low cycle fatigue that has been modeled by the Coffin-Manson equation [1,11]. The interface metallurgy between the solder and the wetting pad is critical to the solder joint reliability. In the present work, we will show that the degraded interface is present due to the high electroplating current density.

The purpose of the current research is to characterize the interface between the solder bump and the wetting pad metallization. The solder is electroplated, and during this operation, the interface experiences resistive heating which leads to interdiffusion and other diffusion controlled reactions, so, characterizing the material system in the as deposited state is very important, and may be used as a reliability indicator.

In the present work, Rutherford Back Scattering(RBS) and X-Ray Diffraction(XRD) have been used to monitor the diffusion and investigate the existence of intermetallic formation respectively. These observations are then used as the basis for process optimization and for reliability prediction.

This study is part of a larger investigation to fully characterize the effect different material parameters have on the evolution of the solder/wetting pad interface. Additional diagnostic tools such as Sputter-Auger spectroscopy and Focused Ion Beam analysis will be used interface analysis. The results will culminate in a physics of failure model for predicting the reliability of the flip chip optical interconnects.

## 2.0 Flip chip design issues

Flip chip design involves material and geometry selection. Both must be selected in order to overcome potential reliability problems [5]. The most significant reliability problem is strain fatigue caused by coefficient of thermal expansion mismatch as mentioned above. It is important to choose materials that have as similar coefficients for thermal expansion as possible. Other problems include solder wetting/dewetting and random fracture at brittle intermetallic layers.

### *2.1 Solder wetting*

The first step in the fabrication of flip chip solder bonds is the formation of the solder wetting the pads. The wetting of solder is controlled by the Young-Dupree Law [12]:

$$\gamma_{sv} - \gamma_{sl} = \gamma_{lv} \cos\theta \qquad (1)$$

Where $\gamma_{xy}$ is the surface energy, and the subscripts refer to interfaces of solid-vapor, solid-liquid, and liquid-vapor. $\theta$ refers to the angle between the solder and the wetting pad. Any factor that effects any of these surface energies, effects the wetting of the solder.[2,11,12]

Oxidation at the solder surface can be a major problem during and after soldering. Even a partial pressure of oxygen of $10^{-6}$torr can be sufficient to oxidize solders [11]. The typical solutions for the oxidation problems include: scrubbing the solder pad before bonding; use of a flux to reduce the surfaces and remove oxides [12]; bonding in an inert atmosphere; and passivation of the bonding surface with a thin gold film. [11]

Another factor that effects the solder wettability is the surface roughness. Shukla and Mencinger have modified from the Young-Dupree equation [2] in order to take into account surface roughness. Equation (2) shows the modified equation in terms of an adhesion energy ($W_a$):

$$-W_a = k\left[\gamma_{lv}\left(1+\cos\theta-\gamma\right)\right] \qquad (2)$$

Where $\gamma_{lv}$ has the usual meaning and $\gamma$ is the strain energy due to thermal mismatch. $k$ is the contact coefficient which depends on surface roughness. This equation has been used to qualitatively compare the surface tensions of the available solders for optical interconnects.

### *2.2 Formation of intermetallics*

In order for a solder bond to adhere to the wetting pad, a metallurgical bond must form [12]. A good indicator of metallurgical bonding is the formation of intermetallics [13] (Intermetallics are chemical compounds of metals with a distinct chemical formula such as $Cu_6Sn_5$ and $Cu_3Sn$).

Without the formation of intermetallics, the joint is only held together by surface tension and Van der Walls forces.

The typical bonding pad metallization in the present investigation includes copper [12], and the wetting pad metallization of Cr/Cu/Au. Solders use a combination of lead and tin, but, environmental concern has generated interest in developing lead free solders. At elevated temperatures, the tin of the solder can diffuse quite readily into the wetting pad. At the interface, a chemical reaction involving the metals of the solder and the metals of the wetting pad can occur forming the intermetallic. Most of these compounds form low melting temperature ternary eutectics. As the ternary eutectic forms, tin can diffuse even faster causing the intermetallic to grow faster. The most important intermetallics and their corresponding eutectics with melting points are listed in table 1.[11]

As has been previously reported, kinetics of intermetallic growth goes through two phases [13]. Initially, the growth of intermetallics is reaction controlled. The intermetallic formation is controlled by the reactivity of the metals. The kinetics then become diffusion controlled since the tin must diffuse to the reacting interface through the intermetallic layer before it can further react with the copper or gold.

### *2.3 Solder fatigue*

A GaAs chip with an area of $1cm^2$ on a silicon substrate will generate approximately 0.08 of shear strain on the flip chip solder bonds with a temperature increase of just 100°C. As the solder is strained, it will elastically and plastically deform. The plastic deformation will result in the material becoming harder and more resistant to further plastic deformation. Further plastic deformation will result in the development of microcracks and eventually fracture. It is possible to design the solder so that the onset of fatigue failure is delayed.

Table 1
Common intermetallics for copper and gold.

| Metal | Intermetallics | Eutectic components | Eutectic Melting Temperature |
|---|---|---|---|
| Gold | $AuSn_4$, $AuSn_2$, AuSn, $AuPb_3$ | Pb+Sn+ $AuSn_4$<br>Pb+Sn+ $AuPb_3$ | 177°C[a]<br>211°C[b] |
| Copper | $Cu_6Sn_5$, $Cu_3Sn$ | Pb+Sn+ $Cu_6Sn_5$ | 182°C[c] |

[a]Prince, Raynor & Evans [14]
[b]Frear, Jones & Kinsman [11]
[c]Chang [15] and Marcotte & Schroeder [16]

Table 2
Fatigue properties for common solders.

| Solder Material | Grain Size | Fatigue Properties ($N_f$)* |
|---|---|---|
| **Hard Solders** | | |
| Au 30% - Sn | 1-5μm | Excellent $>10^4$ |
| Pb 10% - Sn | 5-10μm | Intermediate $10^3$ |
| Pb 10% - Sn | 1-3μm | Excellent $>10^4$ |
| **Soft Solders** | | |
| In | 10-15μm | Poor $10^2$ - $10^3$ |
| PbSn | 1-3μm | Intermediate $10^3$ |
| AuSn | 5-6μm | Intermediate $10^3$ |

Material data taken from ref. 11. Fatigue properties estimated from Coffin-Manson calculations.
* Number of cycles to failure for a ΔT of 125°C

The solder's resistance to fatigue is very dependant on the initial microstructure [1]. If the grain size is sufficiently small, the material will behave superplasitically, and there will be no yield stress observed at equilibrium at either the high or low temperature state of the thermal cycle. However, as the material thermally cycles, the grain size increases reducing its plasticity [7]. The addition of hard nanoparticles such as nickel has been shown to delay the grain coarsening observed in thermal cycling [1]. Table 2 shows the comparisons of solder with respect to microstructure and fatigue resistance for flip chip solder in optical interconnects.

## 3.0 Reliability problems with intermetallics

While the formation of intermetallics is critical for the formation of strong metallurgical bond, excess intermetallic formation reduces the reliability of the solder joint [6,7]. The primary source of failure of solder joints is cracking at the interface between solder and substrate [12]. The

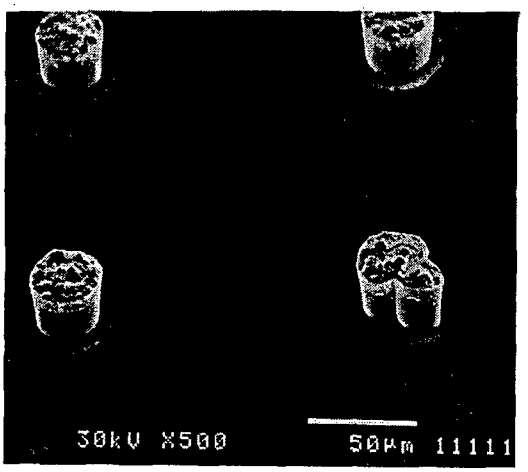



Figure 1. SEM of a sample showing solder bumps

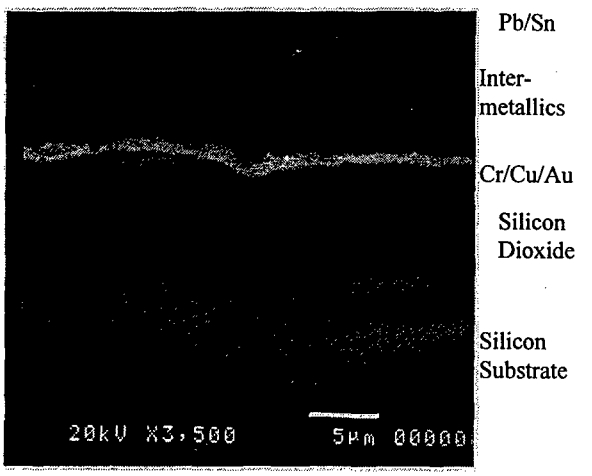



Figure 2. Intermetallics in an as grown sample.

reason for this is that the intermetallics between gold and tin are more brittle than the solder. Gold/tin intermetallics are particularly brittle [11].

A second source of observed failure due to intermetallics is spalling. If there is a large quantity of intermetallic at the interface, it may conglomerate into spheroidal structures that become the source of dewetting during the life of the flip chip solder joint [6].

In order to eliminate these failure mechanisms, one must limit intermetallic formation and inhibit its growth. Keeping process temperatures low during deposition is important for inhibiting growth, while using very thin films of the intermetallic constituents will limit the ultimate thickness of the intermetallic.

## 4.0 Experimental procedures

In the present investigation, samples were prepared by electron beam deposition of 20nm Cr, 150nm Cu and 40nm Au as measured in situ using a crystal deposition monitor. Control samples were analyzed by Rutherford Back Scattering (RBS), while a second set of control samples were analyzed by X-Ray Diffraction (XRD) techniques. These tests were to observe the as deposited condition of the metallization before electroplating.

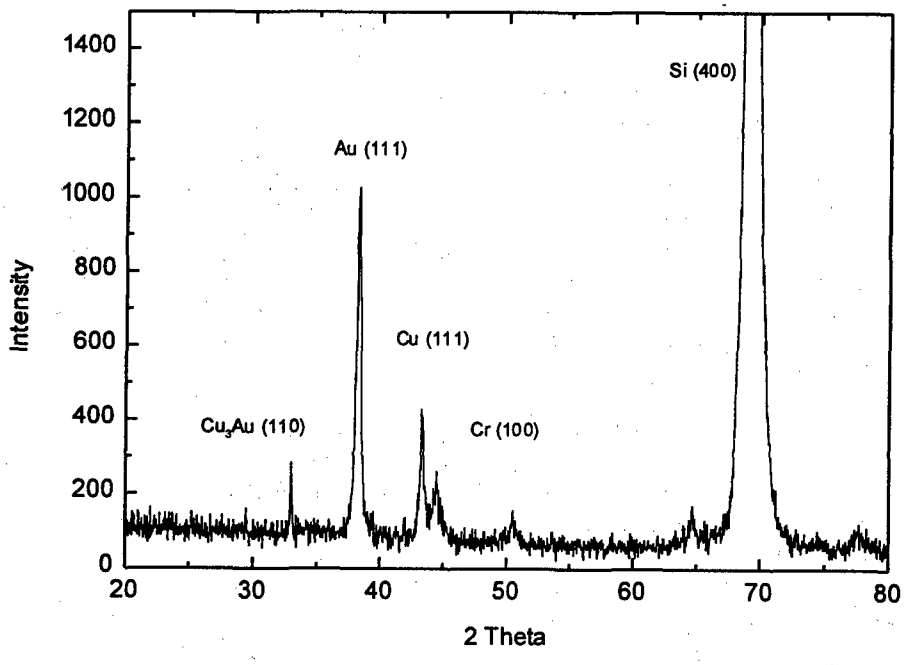
Two sets of samples were prepared using the same metallization on a thermally oxidized silicon substrate. These samples were then electroplated with tin (hard solder) at a current density of 12.4 mA/cm$^2$. The thickness of the electroplated solder was 5µm. The thickness of the tin electroplating was thin enough to allow X-Ray Diffraction (XRD) to be performed. Other cross-sectioned samples were used to measure and monitor the intermetallic thickness. The flip chip samples were 8x8 and 16x16 arrays of solder bumps separated by 256µm.

### *4.1 Experimental results*

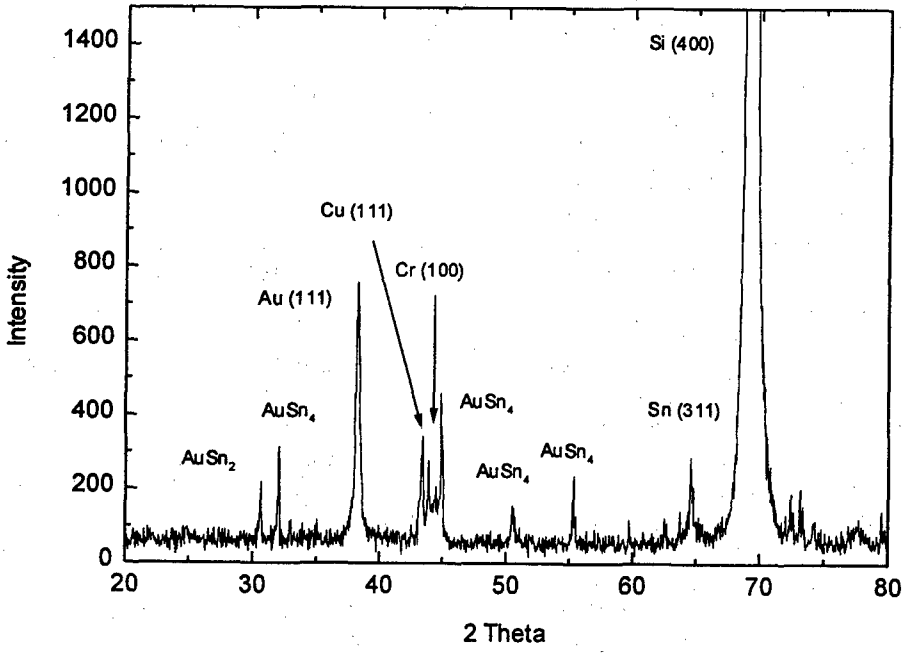
Figures 1 and 2 show scanning electron micrographs of the solder bumps and a cross-section of the structure under analysis. Diffusion of the metallization into the silicon dioxide layer is evident. Also, some round structures are observable in the solder layer. These structures could be intermetallic nodules, or spalling which leads to de-adhesion of the flip chip bond during further processing or testing.

X-Ray Diffraction analysis was also performed on the as deposited Cr/Cu/Au metallization for each type of array as shown in Figure 3(a). The prominent peaks of each metal and the substrate are evident. There is also a peak for a copper—gold intermetallic, showing a well adhered bond between the thin films is present.

Additionally, XRD was performed on the as deposited metallization with the electroplated tin, as shown in Figure 3(b). The analysis was performed prior to reflow or any intentional sintering. The results show Au/Sn intermetallic formation and possible Cu/Sn intermetallic. One can note that the proportion of gold in the $Cu_3Au_2$ intermetallic is reduced. As the gold is consumed at both its upper and lower interfaces, its compounds will tend to be lower in gold content.




(a)




(b)

Figure 3: (a) As grown Cr/Cu/Au on a silicon substrate. (b) Tin electroplated on Cr/Cu/Au metallization. Intermetallic formation is clearly shown.

## 5.0 Conclusion

We have reported the important observation that interdiffusion and compound formation takes place during physical deposition of the materials on the substrate. The interdiffusion of the metals in the wetting pad must take place during deposition since the wafers are kept at room temperature during subsequent processing. The majority of the intermetallic formation between the gold and copper also forms during deposition. The gold--copper intermetallics become more copper rich as the gold is depleted.

Electroplating introduces additional effects that could also have a significant effect on the microstructure. First, there is resistive heating at the interface. This increase in temperature results in enhanced diffusion between the metals and the solder, and increases the intermetallic reaction rate. There is also observable interdiffusion between the metals in the wetting pad and the substrate as well as between the metals themselves. The current during electroplating can also lead to electromigration, which may result in enhanced diffusion in the direction of current flow as well as the formation of voids. Electromigration may also cause localized material buildup resulting in local stress centers.

We conclude that various parameters must be considered when designing solder bonded structures. Geometry and material properties, particularly CTE, must be considered in order to design reliability into the solder bonded structure or module. Material reactions play a very important role in determining the reliability of the module. The most important of these reactions is intermetallic formation that is present in order to ensure a good metallurgical bond. Control of intermetallic growth is critical so that the bond does not become brittle. Coarsening of the solder microstructure must also be addressed in order to ensure reliability during the lifetime of the device. Some other important considerations not discussed in this paper are thermal and environmental management. If the module is subject to cycles of extreme temperature change, the problems associated with thermal cycling will be amplified. If the module is exposed to corrosive environments, the solder may oxidize, or many other reactions could happen to the structure or active components that will be detrimental to device performance. Additional investigations are underway to quantify the observed reactions.

## References:


[1] G. Z. Pan, A. A. Liu, H. K. Kim, K. N. Tu, and P. A. Totta, *Appl. Phys. Lett.* **71**, (20), p. 2946, 17 November 1997

[2] R. K. Shukla and N. P. Mencinger, *Solid State Technology*. p. 67, July 1985

[3] K. L. Lin, and S. Y. Chang, *IEEE Transactions on Components, Packangin, and Manufacturing Technology B* **19**, (4), p. 747, Nov 1996

[4] C. Pusarla, *Design and Fabrication of Hybrid Optical Receiver For Free Space Optical Interconnections Using Flip-Chip Bonding*. Ph. D. Dissertation, Univ. Maryland, College Park. May 1996

[5] S. Greathouse, *Electronic Packaging and Production*. p. 65 August 1997

[6] D. R. Frear and P. T. Vianco, *Metallurgical and Materials Transactions A*, **25A** , p. 1509, July 1994

[7] D. Frear, D. Grivas, and J. W. Morris, Jr. *Journal of Electronic Materials*, **18** (6), p. 671, 1998

[8] J. H. Lau, *Engineering Fracture Mechanics*, **45**, (5), p 643, 1993

[9] A. Yasukawa, *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*. **13**, (4), p. 1146, Dec. 1990

[10] T. Pan, *Journal of Electronic Packaging*, **116**, p. 163, Sept. 1994

[11] D. R. Frear, W. B. Jones, and K. R. Kinsman. *Solder Mechanics: A State of the Art Assessment* Penn. The Minerals, Metals & Materials Society, 1991

[12] F. G. Yost, F. M. Hosking, and D. R. Frear. *The Mechanics of Solder Alloy Wetting & Spreading* New York: Van Nostrand Reinhold, 1993

[13] L. H. Su, Y. W. Yen, C. C. Lin and S. W. Chen, *Metallurgical and Materials Transactions B*. **28B**, p. 927, Oct 1997

[14] A. Prince, G. V. Raynor, and D. S. Evans, *Phase Diagrams of Ternary Gold Alloys*, 1990, Institute of Metals, London.

[15] Y. A. Chang, J. P. neuman, A. Kikula and D. Goldberg. *Phase Diagrams and Thermodynamic Properties of Ternary Copper-Metal Systems*. 1979, INCRA Monographs Series, Vol VI.

[16] V. C. Marcotte, and K. Schroeder. *Proc. 13th No. Amer. Thermal Analysis Soc.* **13**, p511 1984

PERGAMON

Microelectronics Reliability 38 (1998) 1313–1318

MICROELECTRONICS RELIABILITY

# A study of soldering heat evaluation for SMDs

Yoshihiro Etoh, Tatsuo Kayama, Kenji Sasaki

*Quality Assurance Department, Semiconductor Company, Sony Corporation*

**[Abstract]**

Environmental factors affecting package crack occurrence are:
- humid atmosphere and time;
- soldering heat method and the number of times;
- heating temperature and time, etc.

As mounting styles have become diversified, the above environmental conditions have been included in customer's quality requirements. Therefore, reasonable soldering heat evaluation methods are required. This paper demonstrates that resistance to package crack of a package made of a highly adhesive mold resin can be explained in terms of the Interface moisture content and maximum temperature applied to the package only. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Recently, surface mounting devices [hereinafter, called "SMD(s)"] have widely been used because they can be mounted with a high density. Their heating method and the number of times also have become diversified depending on the mounting style. However, sometimes conventional soldering heat evaluation methods are not reasonable or do not always correspond to the diversified mounting style. A reason for the non-correspondence is that many environmental factors give rise to package cracks. [2], [3], [4], [5], [6]

To search for more reasonable evaluation methods, we assessed the effect of the environmental factors again. As a result of the assessment, we have found that the resistance to package crack of biphenyl type epoxy resins which are highly adhesive mold resins is explained in terms of the Interface moisture content and maximum temperature applied to the package only.

## 2. Experiment on Factors Giving Rise to Package Cracks

Factors giving rise to package cracks at customer's warehouse or production line are:
- humid atmosphere and time (which can be considered as one factor, Interface moisture content);
- soldering heat method and the number of times;
- heating temperature (the maximum temperature and profile) and time, etc.

Experiment on the factors giving rise to package cracks are explained below.

### *2.1 Experiment*

Samples and conditions are shown in Tables 2.1.1 and 2.1.2.

Two types of samples whose package outlines are same but are made of different biphenyl type epoxy resins each other were used. Solder dipping method was used for heating. Where heating was performed twice or more, the following procedure was repeated:

1)make samples soak moisture until they are saturated; and then,

2)heat them.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00124-3

**Table 2.1.1: Sample List [Unit: mm]**

| | Sample 1 | Sample 2 |
|---|---|---|
| Package | LQFP-100p | LQFP-100p |
| Resin | A | B |
| Die pad size | 10.2×10.2 | 8.7×8.7 |
| Body size | 14×14×1.4 | 14×14×1.4 |
| Sample size | 10 pcs.×Level 4 | 8 5 pcs.×Level 5 4 |

**Table 2.1.2: Conditions**

| Factor | Level |
|---|---|
| Solder bath temperature | 230°C, 245°C, 260°C |
| Dipping time | 5s, 10s, 30s |
| Number of times | Once, Twice, Four times |
| Interface moisture content | Saturated moisture content at 85°C/65 % RH<br>Saturated moisture content at 85°C/85 % RH |

### *2.2 Results*

The experiment results are shown in Tables 2.2.1 and 2.2.2.
Among the above levels, Interface moisture content and solder bath temperature had striking effects.

The number of times of dipping also had some effect. The reasons are that repeating dipping developed structural damages of the package, such as separation, and caused extra-saturation occurs.

And dipping times of 5s, 10s and 30s had different tendencies. This difference and package temperature profile during the solder dipping indicate the following ;

(1) Cracks occur without depending on the dipping time after the package temperature reaches the solder bath one.

(2) Solder bath temperature and dipping time can be explained as one factor, the maximum package temperature.

## 3.Unifying Solder Dipping Method and Reflow Method

As mentioned above, the maximum temperature of the package is an important factor for solder dipping.

**Table 2.2.1: Sample 1 Results** n=10 for each condition

| ● :With cracks ○ :Without crack | | Saturated moisture content at 85°C/65 % RH | | | Saturated moisture content at 85°C/85 % RH | | |
|---|---|---|---|---|---|---|---|
| | | 230°C | 245°C | 260°C | 230°C | 245°C | 260°C |
| 5s | Once | ○ | ○ | ○ | ○ | ● | ● |
| | Twice | ○ | ○ | ● | ● | ● | — |
| | 4 Times | ○ | ● | ● | ● | ● | — |
| 10s | Once | ○ | ○ | ● | ● | ● | ● |
| | Twice | ○ | ● | ● | ● | ● | — |
| | 4 Times | ○ | ● | ● | ● | ● | — |
| 30s | Once | ○ | ○ | ● | ● | ● | ● |
| | Twice | ○ | ● | ● | ● | ● | — |
| | 4 Times | ○ | ● | ● | ● | ● | — |

**Table 2.2.2: Sample 2 Results** n=10 for each condition

| ● :With cracks ○ :Without crack | | Saturated moisture content at 85°C/65 % RH | | | Saturated moisture content at 85°C/85 % RH | | |
|---|---|---|---|---|---|---|---|
| | | 230°C | 245°C | 260°C | 230°C | 245°C | 260°C |
| 5s | Once | ○ | ○ | ○ | ○ | ● | ● |
| | Twice | ○ | ○ | ○ | ○ | ● | ● |
| | 4 Times | ○ | ○ | ○ | ● | ● | ● |
| 10s | Once | ○ | ○ | ○ | ○ | ● | ● |
| | Twice | ○ | ○ | ○ | ● | ● | ● |
| | 4 Times | ○ | ○ | ○ | ● | ● | ● |
| 30s | Once | ○ | ○ | ○ | ○ | ● | ● |
| | Twice | ○ | ○ | ○ | ○ | ● | ● |
| | 4 Times | ○ | ○ | ○ | ● | ● | ● |

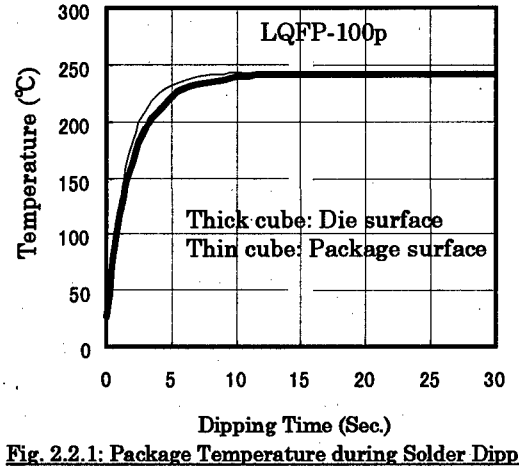



Fig. 2.2.1: Package Temperature during Solder Dipping

By taking notice of this factor, a possibility for unifying two methods of solder dipping and reflow which are defined as different methods each other, is created. We performed an experiment in order to verify the possibility, and the results are as below.

### 3.1 Experiment

Samples and conditions are shown in Tables 3.1.1 and 3.1.2. Two methods were compared in the maximum package surface temperature during the heating is 245℃, the rate of the separated area in the die pad rivers side and whether cracks occurred.

Where heating was performed twice or more, the following procedure was repeated:

1)make samples soak moisture until they are saturated; and then,
2)heat them.

**Table 3.1.1: Sample List [Unit: mm]**

| | Sample 1 | Sample 2 |
|---|---|---|
| Package | LQFP-100p | LQFP-144p |
| Resin | A | B |
| Die pad size | 6.2×6.2 | 9.7×9.7 |
| Body size | 14×14×1.4 | 20×20×1.4 |
| Sample size | 10 pcs.×Level 18 | 5 pcs.×Level 18 |

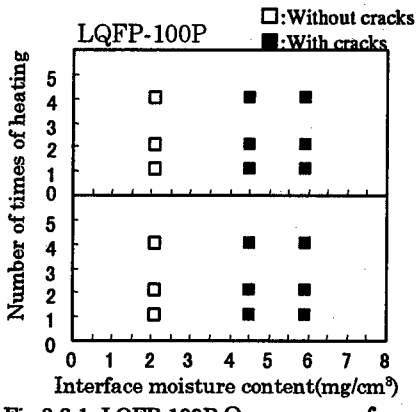



Fig.3.2.1: LQFP-100P Occurrence of cracks

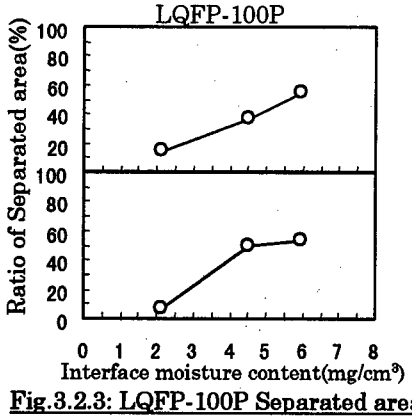



Fig.3.2.3: LQFP-100P Separated area

### 3.2 Results

The results are shown in Figs. 3.2.1 to 3.2.4. Both types of samples exhibited the same tendencies concerning occurrence of cracks and the rate of the separated area in the die pad rivers side regardless of solder dipping or reflow. The results indicate the following:

Although the heating methods are different each other, such as pre-heating is performed or not, the same results can be obtained on condition that they are equal in maximum package temperature.

**Table 3.1.2: Conditions**

| Factor | Level |
|---|---|
| Maximum package temperature | 245℃ |
| Surface moisture content | Saturated moisture content at 85℃/30 %RH<br>Saturated moisture content at 85℃/65 % RH<br>Saturated moisture content at 85℃/85 % RH |
| Number of times of heating | Once, Twice, Four times |

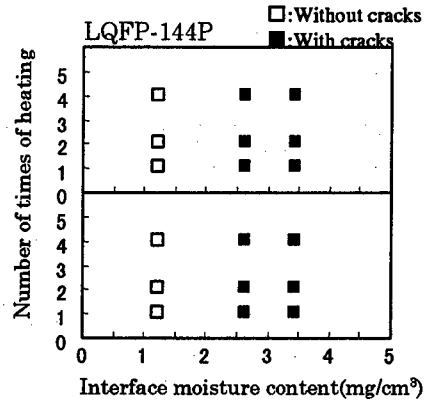



Fig.3.2.2: LQFP-144P Occurrence of cracks

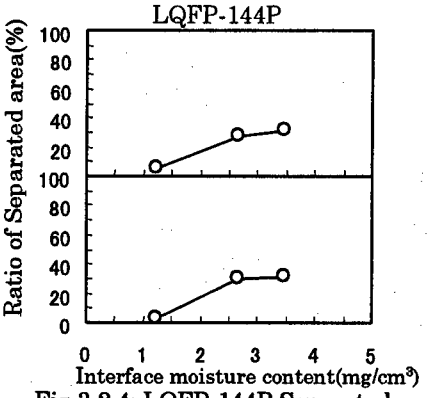



Fig.3.2.4: LQFP-144P Separated area

## 4. Mechanism which Causes Cracks in Highly Adhesive Mold Resins

In general, a package crack occurs when:
-the die or die pad is entirely separated from the resin; and,
-the maximum stress applied to the medium point of the longer leg of the die or die pad, resulting from the vapor pressure at the space between them (P), exceeds the flexural strength $\sigma$ (T).

At this time, the uniform load model for rectangles are used for analysis, and the crack occurrence equation is as follows:

$$\sigma(T) \leqq 6K\left(\frac{a}{h}\right)^2 P \qquad \text{Equation 4.1}$$

a: Length of the die or die pad shorter legs
b: Length of the die or die pad longer legs
h: Resin thickness
k: Constant defined as b/a

However, it is presumed that Equation 4.1 can not always apply to biphenyl type of epoxy resins which have been introduced in order to improve resistance to package crack and are the major package material.

A crack occurrence model for biphenyl type of epoxy resins is explained below.

### *4.1 Crack Occurrence Model*

Adhesion to dice and die pads is the greatest contributor to improvement in resistance to package crack. When a highly adhesive resin is entirely separated from the dice or die pad, the maximum stress $\alpha$ may have already exceeded the flexural strength $\sigma$ (T).

Therefore, at the time of the occurrence of an entire separation, the relation between the vapor pressure P and adhesion per unit is as follows:

$$P > \eta(T) \qquad \text{Equation 4.2}$$

Where saturated vapor pressure in the space: P, saturated moisture density: ds and the moisture density when the vapor pressure is P: d;
P is expressed as follows:

$$P = Ps \cdot \frac{d}{ds} \qquad \text{Equation 4.3}$$

And where the moisture content in the space is assumed to be in proportion to the resin moisture content, complying with Equation 4.3, Equation 4.2 can be converted into

$$Ps \cdot \frac{Q}{Qs} > \eta(T) \qquad \text{Equation 4.4}$$

Qs: Saturated moisture content

Because of temperature (T) of 200 to 300℃, Qs and $\eta$ (T) in equation 4.4 are constant, Qc and $\eta$ c, which is a resin property. And the saturated vapor pressure in the space Ps can be approximated by the following equation:[1]

$$Ps \fallingdotseq B \cdot \exp\left(-\frac{A}{T}\right) \qquad \text{Equation 4.5}$$

A = 4632 [K]
B = 27846 [MPa]

Therefore, Equation 4.4 can be converted into the following by temperature (T) and moisture content (Q):

$$\exp\left(-\frac{A}{T}\right) > Qs \cdot \frac{H(T)}{Q \cdot B} = \frac{Qc \cdot \eta c}{B} \cdot \frac{1}{Q} = \frac{C}{Q} \qquad \text{Equation 4.6}$$

$$\therefore Q > C \exp\left(\frac{A}{T}\right) \qquad \text{Equation 4.7}$$

Constant C depends the on mold resin, die, leadframe construction for good adhesion, and is peculiar to product.

Therefore, Equation 4.7 is the crack occurrence condition for highly adhesive resins, such as biphenyl type epoxy resins. Where one of the 2 parameters, Q or C, is defined, another one is given.

The above is the model we have proposed. If Equation 4.7 is valid, we can estimate the maximum moisture content to be soaked and heating temperature of each package without experiment on condition that each constant C is founded.

### 4.2 Verification

An experiment to verify the crack occurrence condition given by Equation 4.7 is described below:

#### 4.2.1 Details

Samples and conditions are shown in Tables 4.2.1 and 4.2.2.

3 types of packages were used. Solder dipping method were adopted in order to prevent data from dispersing.

The following procedure was repeated 4 times:
1)make samples soak moisture until they are saturated; and then,
2)heat them.

Table 4.2.1: Sample List [Unit: mm]

| | Sample 1 | Sample 2 | Sample 3 |
|---|---|---|---|
| Package | LQFP-100p | TSOP-28p | LQFP-144p |
| Resin | A | A | B |
| Die pad size | 9.7×9.7 | 3.1×6.3 | 5.7×5.7 |
| Body size | 14×14×1.4 | 11.8×8×0.97 | 120×20×1.4 |
| Sample size | 10 pcs.× Level 20 | 10 pcs.× Level 20 | 10 pcs.× Level 20 |

Table 4.2.2: Conditions

| Factor | Level |
|---|---|
| Interface moisture content | Saturated moisture content at 85℃/55 % RH<br>Saturated moisture content at 85℃/65 % RH<br>Saturated moisture content at 85℃/75 % RH<br>Saturated moisture content at 85℃/85 % RH |
| Maximum package temperature | 215℃, 230℃, 245℃, 260℃, 270℃, 290℃ |

#### 4.2.2 Results

The results are shown in Figs. 4-2-1 to 4-2-3. The data are classified into the following and plotted them:

-No crack occurred even after the 4th heating;
-Cracks occurred at the 2nd/ 3rd or 4th heating; and,
-Cracks occurred at the 1st heating.

We found that borders are straight lines regardless of sample type and their slopes are equal to that of Equation 4.5 which approximates saturated moisture content. Therefore, by taking notice of surface moisture content and maximum package surface temperature, the area where cracks would occur given by Equation 4.7.

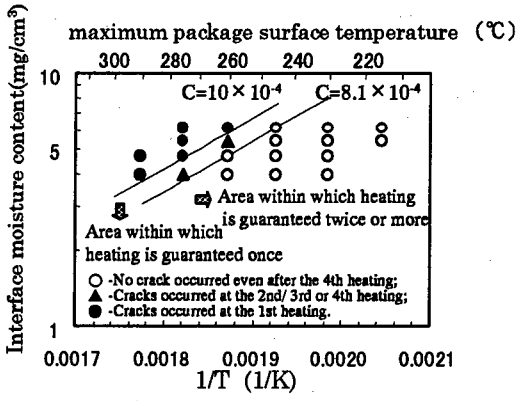



Fig.4.2.1: LQFP-100P Results

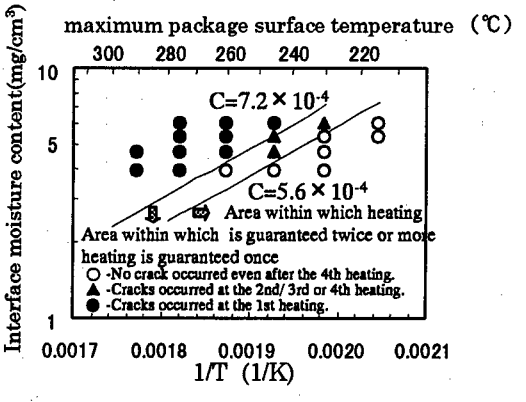



Fig.4.2.2: TSOP-28P Results

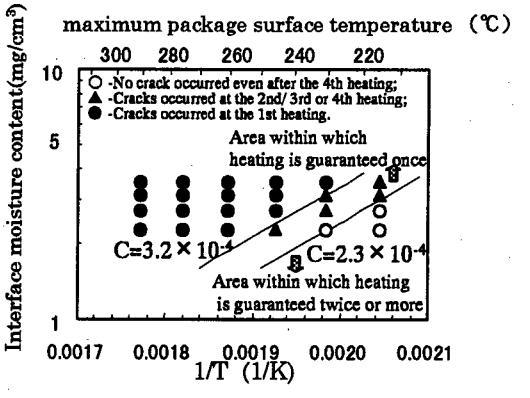



Fig.4.2.3: LQFP-144P Results

Therefore, if the value of constant C has been known, we can presume when package cracks occur corresponding to various mounting conditions in terms of the surface moisture content and maximum package surface temperature.

Values of Constant C of each of the packages used for obtaining soldering heat conditions to be guaranteed are shown in Table 4.2.3:

Table 4.2.3: Soldering Heat Conditions to be Guaranteed and Constants [$\times 10^{-4}$]

| Package Type | Number of Solder Heating to be Guaranteed | |
|---|---|---|
| | Once | Twice or More |
| LQFP-100p | $8.1 \leqq C < 10$ | $C < 8.1$ |
| TSOP-28p | $5.6 \leqq C < 7.2$ | $C < 5.6$ |
| LQFP-144p | $2.3 \leqq C < 3.2$ | $C < 2.3$ |

## 5.Summery

For examination of reasonable evaluating methods of "Resistance to Soldering Heat" we assessed effects of the factors resulting in package cracks again, and the following are found:

(1) By paying attention to the maximum package surface temperature, solder dipping and reflow methods can be unified into soldering heat.
(2)And, resistance to package crack of highly adhesive mold resins can be explained by Interface moisture content and maximum package surface temperature.

We will continue the assessment for trying to apply these models also to other product categories.

## References

[1] "A Study of Reliability Models of HAST" by S. Murakami, K. Syohno and H. Sekiya (Fujitsu Ltd.), at 22nd Reliability and Maintainability Symposium held by the Union of Japanese Scientists and Engineers in 1992

[2] by M. Tanaka, M. Sakimoto, H. Konishi, K. Nishi, K. Ohtsuka and T. Yoshida (Hitachi Ltd.), at 18th Reliability and Maintainability Symposium held by the Union of Japanese Scientists and Engineers in 1998

[3] "Stress Evaluation Method from Repetitive Reflow Soldering on SMD" by M. Ueno, T. Udagawa and T. Setoya (Toshiba Corp.), at 24th Reliability and Maintainability Symposium held by the Union of Japanese Scientists and Engineers in 1994

[4] "Repetitive Reflow Soldering Evaluation Method on SMD" by Y. Ohtsu, M. Ohmori, M. Yanaga and Y. Nagao (NEC Corp.) at 24th Reliability and Maintainability Symposium held by the Union of Japanese Scientists and Engineers in 1994

[5] "Examination on Reflow-resistant for Copper Frame SMD Products" by T. Matsuura, T. Setoya, K. Furuta, Y. Terui, Y. Soekawa, S. Koike and Y. Kishimura (Toshiba Corp.) at 5th Reliability Symposium on Electronics Devices held by Reliability Center for Electronic Components of Japan in 1995

[6]"Cracking Mechanism of SMD during Solder Reflow Process and its Countermeasures" by M. Sugawa and S. Mizugashira (Matsushita Electronics Corp.) at 1st Reliability Symposium on Electronics Devices held by Reliability Center for Electronic Components of Japan in 1991

PERGAMON

Microelectronics Reliability 38 (1998) 1319–1323

MICROELECTRONICS RELIABILITY

# Advanced IGBT modules for railway traction applications: Reliability testing

H. Berg[a], E. Wolfgang[b]

[a] *eupec GmbH, Max-Planck-Straße 5, D-59582 Warstein, Germany*
[b] *Siemens AG, Corporate Technology, Otto-Hahn-Ring 6, D-81730 Munich, Germany*

Abstract

IGBT modules for railway traction applications have to be more intensively tested then those for industrial application. The paper describes the state of the art of already existing reliability test as well as a proposal for accelerated power cycling and temperature cycling tests. These tests are a result of the Brite EuRam research project RAPSDRA (reliability of advanced power semiconductor devices for railway traction application).
© 1998 Published by Elsevier Science Ltd. All rights reserved.

## 1. Introduction

IGBT (insulated gate bipolar transistor) modules have been used for many years for variable-speed drives in industry. In the course of the further development of these modules to handle higher blocking voltages and higher currents, they are also becoming of increasing interest for traction applications. The more rigorous demands made on reliability in traction systems exposed to greater climatic and temperature fluctuations as well as the requirement for 35 years of fault-free operation have made it necessary to re-appraise and suitably adapt the reliability tests used hitherto.

Various research programs have been devoted to this topic in recent years. Among them are the Swiss LESIT program (1993 -1995) [1] and the RAPSDRA (reliability of advanced power semiconductor devices for railway traction applications, 1995 -1998) project [2] within Brite EuRam. In parallel to this, the relevant manufacturers have been working continuously on improving the reliability of the IGBT modules and have developed corresponding technologies.

These efforts focus principally on determining the reliability of the bond and solder joints. Thus it must be shown conclusively that bonding and soldering satisfy the requirements to the same degree as the pressure contacts used hitherto in GTO (gate turn-off thyristor) technology.

This paper describes the state of the art in reliability testing of IGBT modules for traction applications. Section two takes a broad look at the requirements on reliability. The next section then examines accelerated reliability tests and discusses possible accelerating factors. The final section looks at the status achieved so far and at subsequent work.

## 2. Reliability requirements

System considerations made it clear that the failure rate of an IGBT module for traction applications should be at least 100 FIT (1 FIT corresponds to one failure in $10^9$ component hours). To confirm this by experiment, more than 10,000 modules would have to be tested for 1,000 hours. Such an experiment is impracticable for various reasons: the number of high-power modules that the traction market needs annually is in the same order as this figure and the costs for the test would thus be too high by several orders of magnitude.

0026-2714/98/$ – see front matter. © 1998 Published by Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00150-4

**Table 1**
**Standardised reliability tests**

| Reliability test | Proposed test conditions | Existing standards | Estimated testing time |
|---|---|---|---|
| Thermal shock test | $T_{stg\ min}$ = - 40°C<br>$T_{stg\ max}$ = + 125°C<br>$t_{cycl}$ = 2 min, $t_{stg}$ = 2 h | DIN IEC 68,T2-14, test Na<br>DIN 45930, T1 A 4.4.4<br>CECC 50,000, C 4.4.4 | max.<br>100 * (2+2) h = 17 days |
| Environmental test | T = 85°C<br>RH = 85%<br>$V_{CE}$ = 80 V DC<br>$V_{GE}$ = -15 V DC<br>1000 h | amendment to<br>DIN IEC 68, T2-3<br>DIN 45930, T1 A 4.4.4<br>CECC 50,000, C 4.4.4<br>prEN 50125 -1 | 1000 h = 42 days |
| Vibration test | a = 1 g<br>const. v : 15 min<br>Var. v : 1h + 45 min | EN 50155<br>EN 50207<br>IEC 1287 - 1 | 3 * 2 h = 6 h |
| Mechanical shock test | | DIN IEC 1373 | |

It has already been mentioned that the modules, and in particular the joints, are stressed especially by temperature fluctuations between -40 and +125°C. Most problems are caused by the different coefficients of expansion of silicon (3 x $10^{-6}$/K), ceramic (AIN = 4.5 x $10^{-6}$/K) and metal (Cu = 17 x $10^{-6}$/K, Al = 23 x $10^{-6}$/K).

The critical factors affecting the stress on the modules during operation apart from the large temperature range are the fluctuations in operating parameters: in addition to general specifications, that can be derived from the operation of a streetcar, trial runs were carried out in Italy , France, Germany and Switzerland [3] within the framework of the RAPSDRA program, and these provide additional data for the design of the reliability tests.

## 3. Reliability tests

Established procedures for determining and securing the reliability of the products in the field have been available for semiconductor components for many years. They are based on testing the components under extreme conditions of temperature, humidity, temperature cycles or mechanical vibration and impact stresses, on determining fault charts and on deriving acceleration factors for field application.

The traction-system users and manufacturers of power semiconductors working together in the RAPSDRA project are convinced that, in addition to the high temperature reverse blocking test (HTRB test), the following reliability tests that have already been standardised are also well suited to evaluating the reliability of high-performance IGBT modules in traction applications (Table 1): Thermal shock test, environmental test (high humidity, high temperature), mechanical vibration test and mechanical shock test.

However, the large number of fluctuations in operating parameters that can occur during operation over a period of up to 35 years in traction applications are of particular importance for the reliability of the components.

Within the framework of the RAPSDRA project, reliability tests that could provide information about the reliability of the modules in traction applications either singly or in combination were proposed for this purpose. They are shown in Table 2. The test procedures may be used to demonstrate that components can in fact withstand the temperature cycles occurring inside them during a period of 35 years of use without failure.

### *3.1. Reliability of Al wire bond technique*

The operating life of the components is determined by the type and quality of the bonding techniques employed: in these, materials with different coefficients of thermal expansion are bonded together.

The emitter and gate contacts of the semiconductor chips are bonded with the aid of aluminium bonds to the copper contacts on the DCB or the main power supply points. The difference in coefficients of expansion of

**Table 2**
**Proposed reliability tests (RAPSDRA)**

| Reliability test | Proposed test conditions | Standards | Estimated testing time |
|---|---|---|---|
| Power cycling 1 (active) | $T_{min}$ = 55°C<br>ΔT = 50°C, 70°C<br>$I_C = I_{cnom}$, $t_{cycl}$ = 3 sec | --- | 3,000,000 * 3 sec<br>= 104 days |
| Power cycling 2 (active) | $T_{min}$ = 55°C<br>ΔT = 50°C, 70°C<br>$I_C = I_{cnom}$, $t_{cycl}$ = 1 min | --- | 100,000 * 1 min<br>= 70 days |
| Thermal cycling (passive) | $T_{min}$ = 25°C<br>$T_{max}$ = 105°C, 125°C<br>$t_{cycl}$ = 4 min | --- | max.<br>10,000 * 4 min<br>= 28 days |

aluminium and silicon leads to lift-off of the bonds due to temperature fluctuations, as exhaustive investigations have shown [4-14].

However, by selecting improved bond wires as well as using optimised bond processes and protection layers, the reliability of this junction has been so greatly improved in recent years that lift-off of the bonds can be avoided with certainty. Long-term tests have shown that a heel crack caused by bending moments in the Al bond wire leads to failure of the bonded junction only after a significantly increased number of temperature cycles (Fig. 1)

If the tests are carried out with different temperature increases, acceleration factors can be determined that provide information about the reliability of the components for the range of $\Delta T_{junction}$ of 20 - 40 K that is typical for traction applications (Fig. 2).

### *3.2. Reliability of solder layers*

In addition to the bond junctions, the solder layer between the DCB ceramic and the Cu base plate is also exposed to particular stresses by the temperature cycling. This is because the solder layer is subject to mechanical stress due to the different coefficients of thermal expansion of copper and ceramic [12,15].

The result is delamination of the DCB ceramic from the solder layer, a consequent increase in thermal resistance and finally thermal runaway in the affected areas of the module. These effects are independent of whether the temperature cycles are caused by external heating (thermal cycling, Table 2), or by internal heating due to electrical losses in the component (power cycling 2, Table 2). A typical damage pattern for a solder layer is shown by the ultrasonic image in Fig. 3a.

A simple representation of the attainable cycle numbers as a function of $\Delta T_{case}$ (temperature of the base plate), such as that shown in Fig. 2, is not helpful to the user in this case, as the damage here

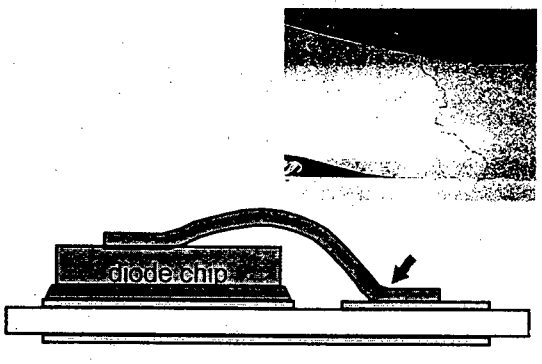



Figure 1: Failure mechanism of heel crack in Al-bond junctions after accelerated power cycling test

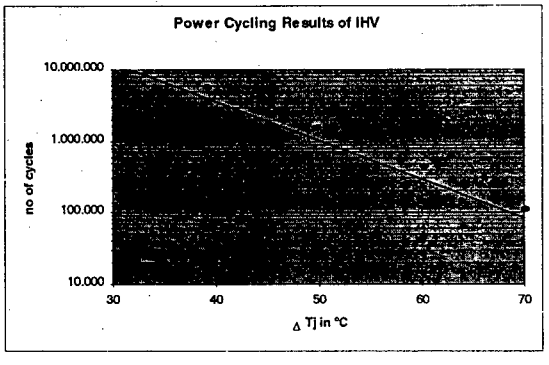



Figure 2: Results of power cycling tests of 1200A/1600V IGBT modules with improved bonding technique for traction applications, $T_{jmax}$ = 100 - 125°C

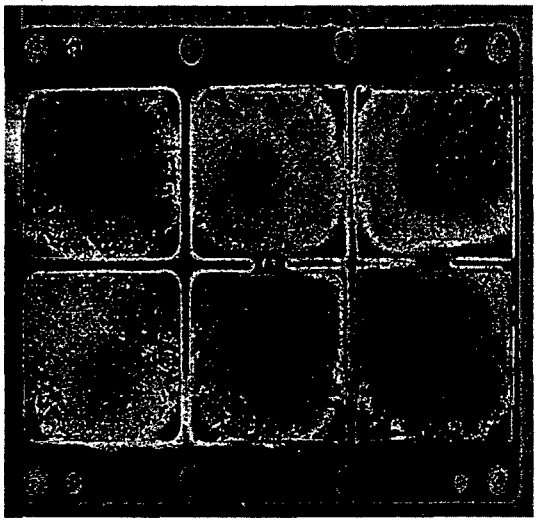


Figure 3a: Solder layer after 2,000 temperature cycles for 1200A/3300V IGBT modules with copper base plate

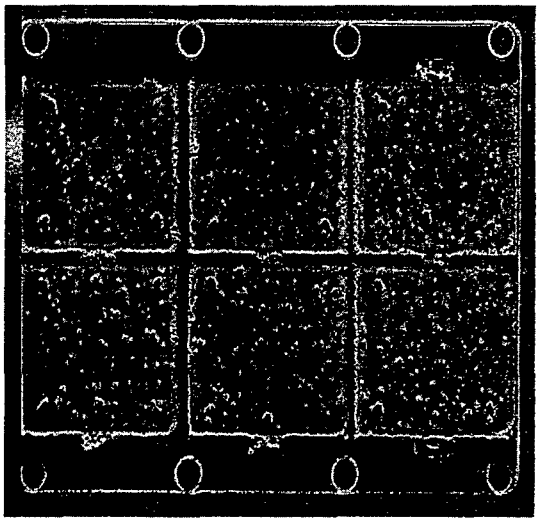


Figure 3b: Solder layer after 20,000 temperature cycles for 1200A/3300V IGBT modules with metal matrix base plate

depends not only on $\Delta T_{case}$ but to an equal extent on the temperature change of the ceramic $\Delta T_{ceramic}$ and the cycle time $t_{cycl}$ in the test. A multidimensional representation is hardly feasible, as the longer cycle times (a few minutes instead of several seconds) and the greater number of test modules means that the tests are no longer practicable for reasons of time and cost.

In the meantime, however, dramatic improvements have been made in module construction, and these lead us to expect that this failure mechanism can be completely avoided in the future. If the Cu base plate is replaced by a material whose coefficient of thermal expansion is better matched to the DCB, the solder will no longer be subjected to mechanical stresses. Suitable materials for this purpose are molybdenum or metal matrix. Figure 3b shows solder layers for modules with metal matrix base plates after ten times as many temperature cycles as for a Cu base plate (Fig. 3a). Although the number of cycles was a whole order of magnitude greater, no damage of the solder layer can yet be detected. It will be the task of the coming months to determine where the limits of this technology lie and whether other failure mechanisms may perhaps be observed at a very much more improved level.

## 4. Discussion

In view of the limited possibility of carrying out reliability tests on thousands of modules, extensive use is made of the concept of built-in reliability [16]. This means that all important parameters affecting the manufacturing process are monitored with extreme precision by means of statistical process control. This applies to both the materials (supplied bonding wires, process gases and liquids) and the process parameters (e.g. temperature). Despite all these measures, however, a limited number of reliability tests must still be carried out.

The intensive work performed on the reliability testing of IGBT modules for traction applications within the Brite EuRam RAPSDRA projects (and previously the LESIT projects) has so far led to the following results:

i) The requirements profiles on the IGBT modules were defined in more detail and were backed up by experimental measurements.

ii) The fault mechanisms were studied intensively by several working groups and a good understanding was obtained of phenomena such as crack propagation in the Al bond wires. This does not originate from electro-migration or stress migration, but from the thermal mismatch between the Al bond wire and the Si chip.

iii) A good understanding of the failure mechanisms occurring in the tests and in the field makes it considerably easier for the process developers to introduce technological improvements. This was successfully done outside the project by all RAPSDRA partners.

iv) The RAPSDRA consortium has proposed test conditions for two power-cycling tests and a thermal cycling test in addition to the already standardised reliability tests such as the thermal shock test, the environmental test, the vibration test and the mechanical shock test. The first power-cycling test, that focuses on the bond wire reliability, could be based on earlier work performed within the scope of the LESIT project. The second power cycling test and the thermal cycling test will investigate the solder reliability. However, final test conditions still remain to be defined and examined within the RAPSDRA consortium.

It is the avowed aim of RAPSDRA to propose the elaborated reliability tests that have been co-ordinated between the manufacturers of IGBT modules and converters on the one hand and the railway operators on the other hand to CENELEC as standardised tests. As IGBT module technology is currently in the throes of extremely rapid development, the RAPSDRA consortium plans to create a special working group within CENELEC TC9X that will be dedicated to the topic of reliability testing for railway traction applications.

## Acknowledgements

We would like to express our gratitude for the extremely good and constructive co-operation in the RAPSDRA working group on Task 3, accelerated reliability testing. In particular, we wish to thank J. Coquery, INRETS, Prof. F. Fantini, Univ. Parma, D. Newcombe, Mitel Semiconductor, Dr. R. Zehringer, ABB, G.P. Sanguinetti, Ansaldo.

## References

[1] LESIT (Leistungelektronik; Systemtechnik, Informationstechnologie): Swiss Federal Research Project (1993 - 1995)

[2] Brite EuRam BE 95-2105 RAPSDRA: Reliability of Advanced Power Semiconductors for Railway Traction Applications

[3] Zehringer, R.; Struck, A.; Lange, T.: Material Requirements for High Voltage, High Power IGBT devices, MRS Symposium Proceedings Nol. 483, (1998), pp. 369 – 380

[4] Jacob, P.; Held, M.; Scacco, P.; Wu, W.: Reliability Testing and Analysis of IGBT-Power Semiconductor Modules, Proc. 20st ISTFA, (1994)

[5] Malberti, P.; Ciappa, M.; Cattomio, R.: A Power-Cycling-Induced Failure Mechanism of IGBT Multichip Modules

[6] Wu, W.; Held, M.; Jacob, P.; Scacco, P.; Birolini, A.: Investigation on the Long Term Reliability of Power IGBT Modules, Proc. ISPSD 95 (1995), pp. 443 - 448

[7] Poech, M.-H..; Dittmer, K.J.; Gäbisch, D.: Investigations on the Damage Mechanism of Aluminium Wire Bonds Used for High Power Applications, Proc. EUPAC 96 (1996), pp. 128 - 131

[8] Held, M.; Jacob, P.; Nicoletti, G.; Scacco, P.; Poech, M.-H.: Fast Power Cycling Test for IGBT Modules in Traction Application, Proc. PEDS 97 (1997), pp. 425 - 430

[9] Auerbach, F.: Power-Cycling-Stability of IGBT-Modules. Proc. IEEE IAS, New Orleans, October 5-9 (1997)

[10] Sommer, K.H.; Spanke, R.; Lefranc, G.; Göttert, J.: Multichip High Power IGBT-Modules for Traction and Industrial Application, Proc. EPE' 97, Trondheim, Norway, September 22-24, pp. 512-515

[11] Hamidi, A.; Coquery, G.; Lallemand, R.: Reliability of High Power IGBT- Modules Testing on Thermal Fatigue Effects due to Traction Cycles: Proc. EPE '97, Trondheim

[12] Hamidi, A.; Coquery, G.; Lallemand, R.; Vales, P.; Dorkel, J.M.: Temperature Measurements and Thermal Modeling of High Power IGBT Multichip Modules for Reliability Investigations in Traction Applications (this proceedings)

[13] Cova, P.; Fantini, F.: On the Effect of Power Cycling Stress on IGBT Modules (this proceedings)

[14] Ramminger, S.; Türkes, P.; Wachuta, G.; Crack Mechanism in Wire Bonding Joints (this proceedings)

[15] Herr, E.; Frey, T.; Schlegel, R.; Stuck, A.; Zehringer, R.: Substrate-To-Base Solder Joint Reliability in High Power IGBT Modules, Microelectronics and Reliability, Special Issue ESREF 97 (1997), pp. 1719 - 1722

[16] Crook, D.L.: Evolution of VLSI Reliability Engineering. Proc. of IRPS 1990, pp. 2-11
Brite EuRam BE 95-2105 RAPSDRA: Reliability of Advanced Power Semiconductors for Railway Traction Applications

PERGAMON

Microelectronics Reliability 38 (1998) 1325–1330

MICROELECTRONICS RELIABILITY

# Some observation dealing with the failures of IGBT transistors in high power converters

**S. Januszewski, M. Kociszewska–Szczerbik, H. Świątek**

*Electrotechnical Institute, Power Converter Department,*
Pożaryskiego 28, 04–703 Warszawa, Poland

**Abstract**

The explosion strength of high power IGBT modules is one of the important parameters that may decide on converter equipment reliability in extreme circumstances. The explosion strength of a device is represented by the peak value of the collector current (so called „peak case nonrupture current") that cannot be exceeded. Some main IGBT failure mechanisms under service conditions are discussed. Test methods and experimental results on device case explosion strength are presented. Experiences concerning starting, tests and service of high IGBT converters are given. Preliminary load tests with rapidly changing current values are useful for newly fabricated IGBT converters. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Most papers at international symposiums and conferences dealing with power semiconductor devices (PSD) and their applications are connected with theoretical, technological and development consideration in these fields. Semiconductor converters have been in the industrial service since more than 30 years with a positive experience referring reliability and lifetime.

But now this situation may be changed. There will be more and more high voltage and high power electronics installations [1,2]. In such converters controlled power semiconductor devices of new generation are used, especially IGBT transistors. The service reliability of these installations is not known yet. Actual experiences dealing with few years observations of old generation IGBT converters are rather positive but a few years of service time are not sufficient. New generation IGBT modules with a trench gate construction are implemented now in power electronics equipment.

## 2. Characterisation of the structure of IGBT modules

Some microelectronics technological processes are successfully used in modern power semiconductor device manufacture. Geometry and surface particle density are dictated by the linewidth requirement (Fig.1). Segment size is gradually decreased. But an increase in IGBT voltage

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00091-2

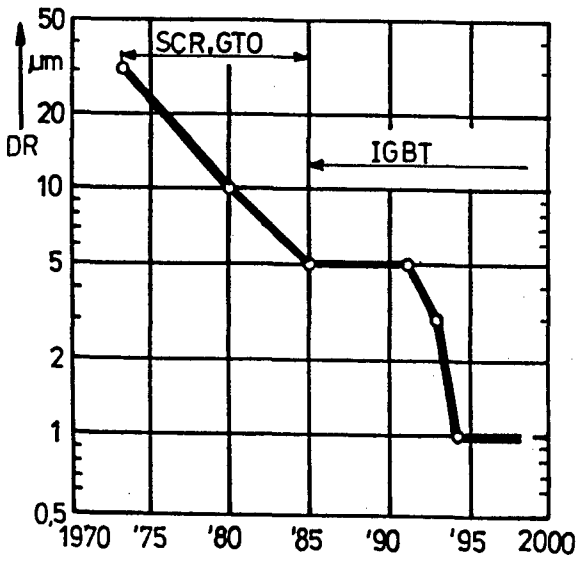



Fig.1 Linewidth (Design Rule - DR) development of IGBT power transistors in comparison to SCR and GTO thyristors

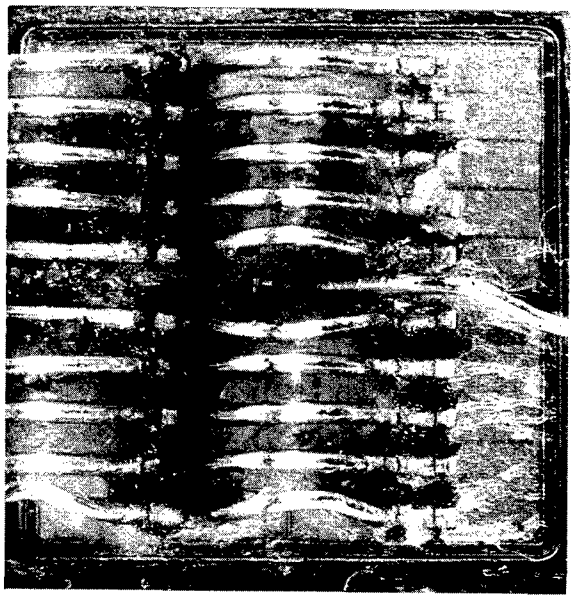


Fig.2 A fragment (one chip) of the bottom part of destroyed IGBT module (400A, 1200V). The silicon gel was removed. All bonding wires are either fused or lifted-open (dark colour)

ratings connected with the decrease in segment size create new unknown reliability problems growing under severe load conditions in high voltage and high power IGBT converters.

A power IGBT module is a multichip device (Fig.2). Therefore a current unbalance among IGBT chips, due to the differences of electrical parameters such as threshold voltage $U_{GE(th)}$, on state voltage $U_{CE(on)}$, and switching times $t_{on}$ and $t_{off}$, might cause thermal failure of IGBT. Most of IGBT chips of the failed modules are burned out either catastrophically or locally. All the emitter and gate bonding wires are either fused or lifted–open.

An IGBT module has a multilayer structure in its package. When it is subjected to temperature changes like in the load cycling, thermal stress due to the difference of thermal expansion coefficient of each material in the structure will be developed (bimetal effect), which may result in a fatigue failure in soft solder and make silicon chip break because of its brittleness. The peak value of load current and the maximum structure temperature appear very clearly as important factors of reduction of life time for big IGBT modules.

Wire bonding is usually considered as one of the weakest areas of device packaging, which is especially true for power IGBT modules because thick wires of 0,25 - 0,5 mm in diameter are used in IGBT to conduct high current. Bonding pressure and temperature are two sensitive factors determining bonding quality. With a poor bonding pressure it is impossible to form a good contact between wire and bonding pad. This fact may lead to a low fracture strength at the bonding interface. Moreover, bonding wires are subjected to a tensile stress due to the changes of temperature during power cycling.

Surface degradation in emitter bonding pads was observed when seeking for the reason of wire lifting failure. With power cycling the metallization surface became rough and hillocks were formed because of recrystallization and electromigration of Al metallization. This would be accelerated by thermomechanical stress due to the difference in thermal expansion coefficients of Al and Si.

The frequency of heavy surge current appearance has great influence on long term IGBT converter reliability. Each following overload influences the reliability of large area die bonding of power IGBT modules by the thermomechanical stress behaviour. This kind of service is not preferred for power IGBT devices. The forming and growing process of voids and cracks in the IGBT solder layer was observed during thermal cycling tests. Furthermore, it is evident that the voids and cracks caused by thermal stress will degrade heat dissipation of IGBT modules. All these circumstances have negative influence on IGBT modules with single side cooling.

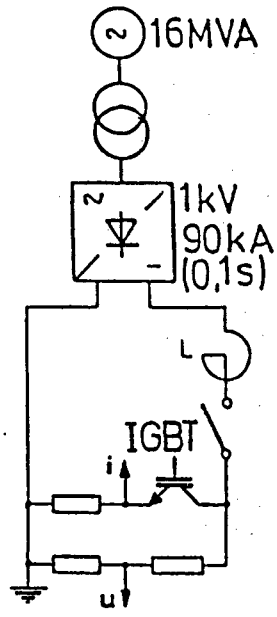



Fig.3 The IGBT module (400A, 1200V): Basic d.c. test circuit used during short-circuit test. Allowable values of short-circuit current and time duration are given in brackets

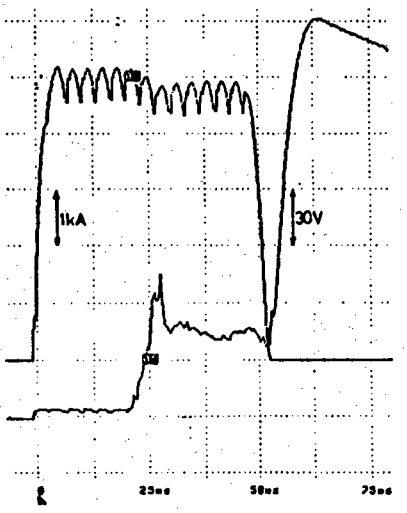



Fig.4 The IGBT module (400A,1200V): Voltage and current waveforms recorded during the case explosion

## 3. The peak case non-rupture current of IGBT modules

The explosion strength of IGBT modules is one of the important parameters that may decide on converter equipment reliability under extreme work conditions. This explosion strength is represented by the peak value of the main current (so called „peak case non-rupture current") that must not be exceeded to prevent dehermetization of the device case.

During converter equipment operation the following phenomena were observed (they lead to dehermetization of power semiconductor devices):

- gradual loss the sealing of some device case due to undetected production defects that shown themselves during device service, particularly in variable load conditions,
- dehermetization of power semiconductor device case due to improper procedure during maintenance, for example a strain crack of enclosure caused by incorrect replacement of failed device,
- dehermetization of power device case (sometimes in the explosion mode) due to the high value of the short-circuit current after device breakdown.

From the service point of view a case explosion is the worst case. The arc phenomena are a significant part of all electrical damages. They are the more dangerous the greater the arc current and the longer the time of its flow. Serious failures may appear in the equipment already after very short time (few milliseconds).

In the case of an internal short-circuit caused by the breakdown of semiconductor devices a high current (50÷100kA) can flow through the damaged device. This is especially dangerous in the case of higher DC voltages (more than 500V). Dehermetization of a semiconductor device case can occur before the action of overcurrent protection under discussed circumstances. Then a plasma stream is emitted to the environment. In the condition of high short-circuit current the device case can explode even before the action of fuses. Then the damage is propagated throughout converter cubicles, which can eliminate them from operation (destructive action of electrical arc).

## 4. Investigation of IGBT module explosion strength

An international IEC standard dealing with uniform test methods of IGBT transistor parameters is actually under consideration within the IEC Technical Committee TC 47.

In the Electrotechnical Institute (Warsaw) investigations dealing with the power semiconductor device explosion strength have been carried out for many years. The aim of these tests has been to obtain an information necessary for rational

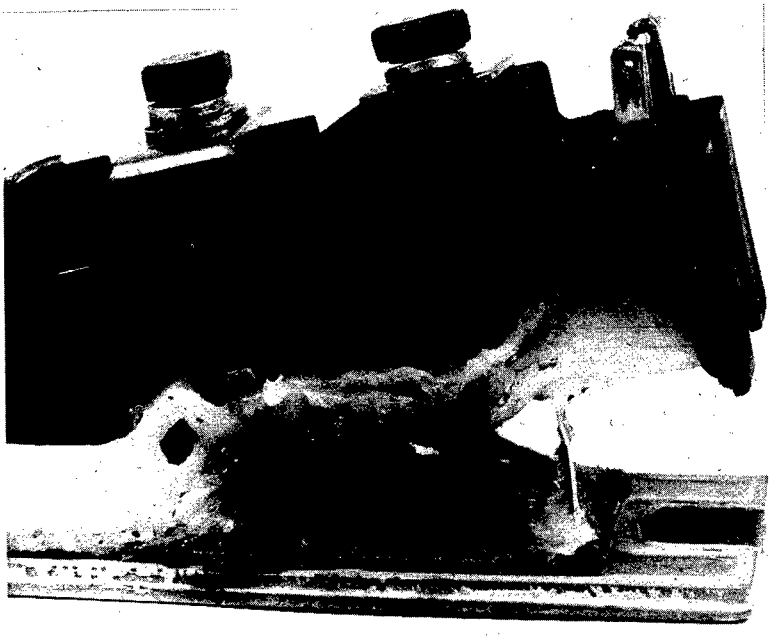


Fig. 5 IGBT module (75A, 600V) composed of two transistor and two antiparallel fast diodes. It was exploded by short-circuit current of about 5kA (peak value) after 6-7ms. Lateral parts of the case were thrown aside and the top part was lifted from the bottom one. During test an electric arc was observed on both lateral sides of the module. White silicone gel is visible between both destroyed fragments of the case. A central internal part of the gel was melted and it solidified in the form of a black layer bloom on the surface of wire bonding.

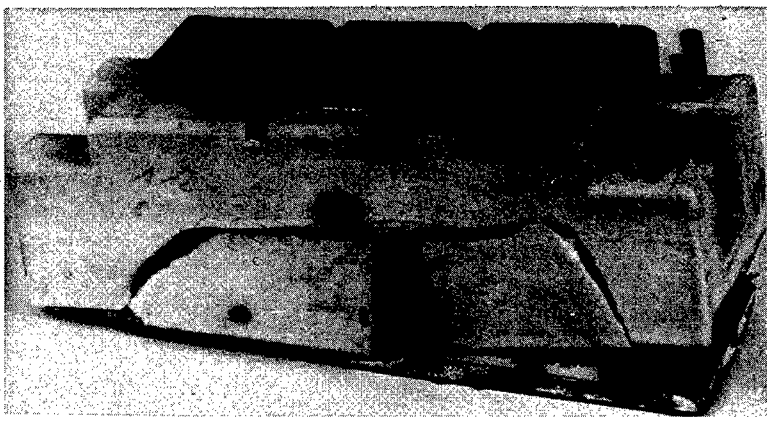


Fig. 6 IGBT module (75A, 600V) connected in half bridge circuit. It was exploded by short-circuit current value of about 5kA (peak value) after 6ms. Lateral parts of the module were thrown open and the top part was lifted from bottom one.

design of overcurrent protection system for very high power semiconductor converter equipment.

Empirical investigations of an IGBT module explosion strength were carried out in the test circuits shown in Fig.3 under normal service conditions. Tested IGBTs were destroyed during first laboratory and service experiments with new designed high power converters. Blocking voltage capabilities of these devices were lost at that time. During explosion tests a current can easily flow through the failed IGBT. The gate circuit was open. Short-circuit waveforms of currents and voltage, connected with explosion phenomena are recorded (Fig.4). Multivariant experiments of the influence of increased peak pulse current value and their duration time on the explosion strength of various kind of IGBT modules were performed. The destruction test method has been applied. Some exemplary photos of modules destroyed by explosion are given in Fig.5 and 6 with appropriate comments. Explosion strength test results of 400A modules are presented in Fig.7.

Observations dealing with the explosion of IGBT module plastic package show that almost always the top part of the case was lifted from the bottom one connected with a sink area (Fig.5). As a rule two sides of thin lateral parts of this module are thrown open as well. During the explosion an electric is formed. It persists till the gas pressure creates a crack in the module and then a plasma stream is emitted to the environment. Most of the internal construction details gets melted by the electric arc and thrown away. The internal silicone gel is melted as well and it solidifies in the form of black layer bloom on the surface.

Phenomena related to the IGBT module explosion appear more intensively in the case of high voltage converters supplied from high short circuit power capacity systems.

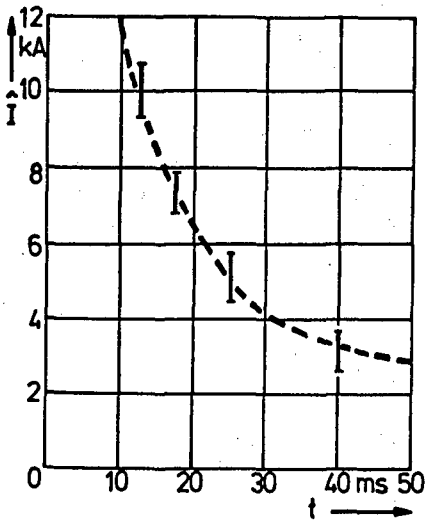



Fig. 7 Peak case rupture current of IGBT modules (400A, 1,2kV) vs. time (see Fig.3 and 4)

Development of a new power IGBT converter is almost always connected with the possibility of sporadic device failures. Each IGBT module used in power electronics equipment is at first subjected to a simultaneous current-voltage stress in normal service conditions, taking into account some different interactions. During routine type tests of a device in a factory such circumstances do not exist.

## 5. Internal short–circuits in IGBT converters

High power IGBT converters, as a rule, are designed as three phase bridge circuit connections. The most severe service conditions in this topology are internal short-circuits caused by semiconductor device breakdown. In such situation a high fault current will cause the explosion of an IGBT because of the energy built up inside the transistor.

In the case of the IGBT high voltage source inverter, often used in a.c. electrical drive systems (Fig.8), the energy accumulated in the capacitor C is very dangerous from the short circuit point of view. Some malfunctions in IGBT control circuits and diagnostic systems of the same phase two arms or simultaneous breakdown of these both transistors cause very severe short-circuit with great probability of case explosion and dangerous consequences for the whole converter (a possibility of fire).

A high fault current may cause the explosion of the IGBT (as for other PSD) because of the energy built up inside the component [5,7].

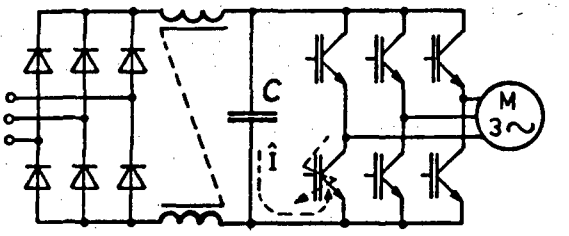



Fig.8 Main circuit of high power three phase bridge indirect converter with IGBT voltage source inverter. Fast free-wheeling diodes connected in antiparallel with the transistors are omitted

The short circuit peak current value can reach several tens of kiloampers and the IGBT case explosion occurs very quickly if there is no current limiting fuse in the circuit. Fast fuses must be selected to prevent any damage to the IGBT case because the IGBT structure cannot be protected by fuses due to the IGBT extremely low $I^2t$ value.

Majority of IGBT module failures in the service conditions take place due to gradual lifting of the bonding in internal device construction. The IGBT modules in plastic cases, have low explosion strength as compared to disc semiconductor assembled with press–pack technology. Short–circuit currents with values as low as several kiloampers cause the case explosion of these IGBT devices. Introduction of a disc shape construction for high power IGBT enables two side cooling and limits consequences of explosions similarly as in the high power diodes and thyristors of such design [8].

## 6. Conclusions

The explosion strength of high power IGBT module packages must be specified for full quality assessment of these devices. Majority of IGBT module failures in service conditions take place due to gradual lifting of the bonding in internal device construction. These wire bondings are the weakest parts of the high power IGBT module with one side mode of cooling.

IGBT modules in plastic cases, have small explosion strength when compared to disc semicon-

ductor assembled with press-pack technology. Short-circuit currents of values as low as several kiloampers cause case explosion of these IGBT devices. Introduction of a disc shape construction with press-pack technology for high power IGBT [8] enables two side cooling and limits consequences of explosions [8] similarly as in the high power diodes and thyristors with such design.

It is useful to introduce more than twenty hours preliminary load tests with rapidly changing current values for new fabricated high power IGBT converters [6]. This will significantly diminish the number of potential failures of IGBT's that would be very dangerous from the explosion point of view in normal service conditions. The fire hazard to the whole power electronic equipment will be lower as well. The application of correctly selected fast fuses [3,7] in high voltage IGBT converters is a necessity from the technical and economic points of view. It will effectively protect the equipment against destruction of converter internal parts when the IGBT module explodes.

## References


1. Brunner H. et al.: 3300V IGBT module for traction application. EPE'95, Seville, vol.1, p. 56-59
2. Coquery G. et al.: Reliability of the 400A IGBT modules for traction converters - contribution on the power thermal fatigue influence on life expectancy. EPE'95, Seville, vol.1, 60-65
3. Duong s. et al.: Short circuit behaviour for PT and NPT IGBT devices – protection against explosion of the case by fuses. EPE'95, Seville, vol.1, 249-254.
4. Januszewski S., Kociszewska-Szczerbik M., Stypułkowska E., Świątek H., Świątek G.: Investigation of destroyed parts of surface of high power semiconductor devices in service conditions. 6th European Symposium Reliability of Electron Devices, Failure Physics and Analysis, ESREF'95, Oct. 1995 Bordeaux
5. Januszewski S., Kociszewska-Szczerbik M., Świątek H., Zymmer K.: IGBT transistor failures in high power converters, Proceedings of the 3th International Seminar on Power Semiconductors-ISPS'96, Sept. 1996 Prague (Czech Republic).
6. Januszewski S., Świątek H., Zymmer K.: Some experiences concerning the starting and service of high power IGBT converters, EPE'97 Trondheim, vol.3, 103-106
7. Mulertt C et al.: Fuses against IGBT inverter explosion. PEMC'96 Budapest, vol. 1, 80-87.
8. Takahashi Y. et al.: 2,5kV -1000A power pack IGBT (High power flat-packaged RC-IGBT). International Symposium on Power Semiconductor Devices and ICs - ISPSD'96, May 1996, Hawaii (USA), 299-302.
9. Wu W. et al.: Thermal stress related to packaging failure in power IGBT modules. ISPSD'95, May 1995, Yokohama (Japan), 330-334

PERGAMON

Microelectronics Reliability 38 (1998) 1331–1334

MICROELECTRONICS RELIABILITY

# Strain depending reliability of automotive diodes

L. Galateanu[a], M. G. Stoica[a], E. Popa[b]

[a] *National Institute for Microtechnologies-IMT, 32B Erou Iancu Nicolae Street, Bucharest, Romania*
[b] *S.C. Baneasa S.A.,32 Erou Iancu Nicolae Street, Bucharest, Romania*

**Abstract**

After improving the assembly system, the remaining automotive diodes failures, above 4% in operation, was associated to the internal strain, based on the RVT test results. The X-ray diffraction spectra - the rocking curves - allowed the pointing out of the internal strain induced by the grinding and diffusion processes. New conditions for the diffusion process, which reduce the induced strain on the B-Al diffused side, were established. An important improvement of the automotive diodes reliability was obtained, the failure percentages diminishing by a factor between 4 and 8, in operation and in RVT tests.
© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

The automotive rectifier diodes of $I_F$=25A and $U_R$=200V presented failure percentages above 4% in operation. The failures consisted in strong degradations of the reverse characteristics. SEM analysis pointed out microcracks at the edge of the semiconductor chips. In order to identify the stress responsible for the failures and the weak point of the devices, reliability tests were performed [1]. No failure appeared after vibration and mechanical shock tests, high temperature (+175 $^0$C) and low temperature (-55 $^0$C) storage and high temperature reverse biasing ($T_j$ = +175 $^0$C, $U_R$ = 150V). Only after performing the rapid variations of temperature, the RVT test (15 cycles of 30 min. at extreme temperatures of +150 $^0$C and -55 $^0$), the same failure mode at the same percentages as in operational conditions were obtained. Taking in account that this study was made after the introduction in the assembly system of a new stress relief shape for the internal connection wire, which led to an important reliability improvement [2], the hypothesis that an internal residual strain is responsible for the remaining automotive diodes failures, was advanced.

The X-ray diffraction spectra - the rocking curves - obtained by the double crystal method [3] were used for the study of the residual strain of the ground and diffused wafers. An improved angular resolution of $4.5 \times 10^{-3}$ degrees for the X-ray diffraction spectra was available.

In the paper, the authors present the results obtained in emphasising, by X-ray diffraction spectra, the internal strain induced by grinding and diffusion processes, in diminishing these strains and in improving, in this way, the automotive diodes reliability.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00133-4

## 2. The X-ray diffraction spectra

The X-ray diffraction spectra were obtained by the double crystal technique, using a $CuK_\alpha$ radiation, with P = 40kV x 20mA. The angular scanning was performed in the area corresponding to the X-ray diffraction on the (333) reflex plane with an angular step of 0.0025 $^0$ and an exposure time of 10sec.

The samples were ground and diffused wafers, which present a surface process-strained layer. The internal residual strain modifies the lattice parameters for the surface layer and consequently, the X-ray diffraction spectra will be cosequently larger. Also, the X-ray rocking curve may appear as a two peaks curve, one peak corresponding to the diffraction on the strained surface layer and one peak to the diffraction on the bulk. The width of the rocking curve or the angular distance between the two peaks is a first qualitative measure of the residual strain in the wafer.

For quantitative determinations, the X-ray rocking curves, obtained for the processed wafers, were approximated by the sum of two gaussian profiles. The narrow profile characterises the diffraction on the bulk, while the X-ray scattering on the surface strained layer gives the broader one.

The following parameters, which point out the magnitude of the residual strain in the processed wafer, were calculated:

- *FWHM* - the full width at half maximum,
- $m_\perp$ - the relative elastic deformation, of the lattice, normal to surface,
- $\sigma$ - the internal residual strain.

The relative elastic deformation, normal to surface, $m_\perp$, is defined by the relation:

$$m_\perp = \Delta d/d \qquad (1)$$

where:

- d, $\Delta d$ are the distance between the lattice planes and its variation.

For obtaining the relative lattice deformation, $m_\perp$, the following equation is used:

$$m_\perp = \Delta\theta \, \mathrm{ctg}\, \theta_B \, / \cos^2\phi \qquad (2)$$

where:

- $\Delta\theta$ is the angular distance between the diffraction peaks,
- $\theta_B$ is the Bragg angle,
- $\phi$ is the angle between the wafer surface and the reflex plane.

The internal residual strain, $\sigma$, is given by the equation:

$$\sigma = E(m_\perp - m) / 2\nu \qquad (3)$$

where:

- E is the Young module,
- $\nu$ is the Poisson coefficient,
- m is the lattice misfit factor.

## 3. Grinding induced strain

For reasons pertaining to surface wettability the automotive diodes are made using ground wafers. The mechanical resistance of the ground wafers is lower than of the polished ones. The grinding process leaves behind a damaged layer which can be viewed as consisting of two overlapped areas. The upper one is represented by the so-called roughness and by microcraks, which advance towards the bulk. The second one is the area of an internal strain, which slowly diminishes by depth.

In a previous study [2], the authors showed that the breaking force of the wafer is not always correlated with the roughness depth and the IR absorption at the lowest wavelength (2,5 μm) seems to be depending only upon the surface fine perturbations.

The internal residual strain of the wafer, induced by the grinding process, was pointed out by X- ray diffraction spectra obtained by the method described in chapter 1. The width of the rocking curves stands for the magnitude of the lattice parameter modification, which is determined by the internal strain.

The rocking curves were performed for ground wafers (figure 1a and 1b) and for polished wafers (figure 1c). The comparison emphasises significant greater strains in the ground wafers.

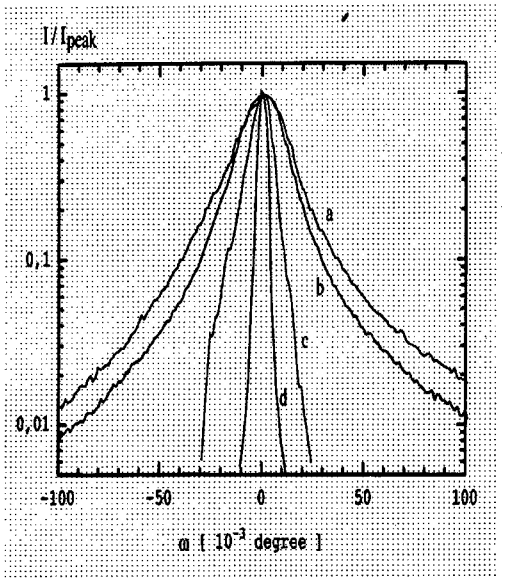



Fig. 1.X-ray diffraction spectra
a- after "18μm" grinding;
b- after "8μm" grinding;
c- after alkali etching;
d- instrumental profile.

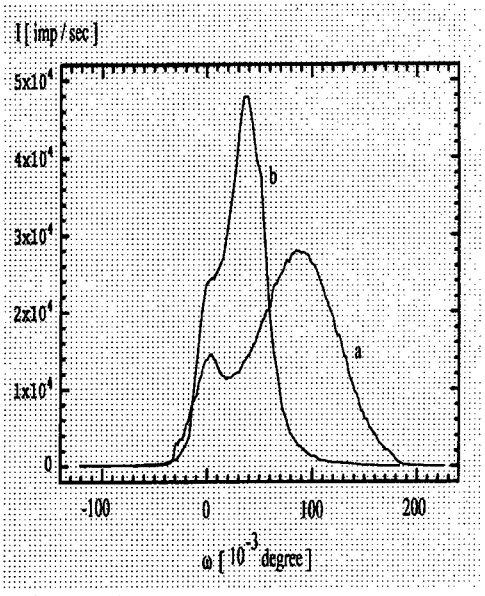



Fig. 2. X-ray diffraction spectra for B-Al diffused side
a- after the standard diffusion process
b- after the improved diffusion process

In order to obtain a lower stressing grinding process, a smaller grain size of the $Al_2O_3$ abrasive was used. The results for the step between 18μm and 8μm of $Al_2O_3$ grain size were analysed. The rocking curve for the grinding process by 8μm grain size (figure 1b) is few narrower than the one for the grinding process by 18μm grain size (figure 1a).

For obtaining quantitative determinations, the rocking curves were approximated by the two gaussian profile components - one corresponding to the surface strained layer and one to the bulk. The $FWHM_{surf}$ and $FWHM_{bulk}$ were calculated for the both grinding conditions. The following results were obtained:

i) grinding by 18μm grain size:

$FWHM_{surf}$= 331″
$FWHM_{bulk}$= 88″

ii) grinding by 8μm grain size:

$FWHM_{surf}$=286″
$FWHM_{bulk}$= 88″.

Therefore, no significant reduction of the internal residual strain was obtained for the grinding process.

## 4. Diffusion induced strain

For the automotive diodes, a double diffusion process is used. B-Al and P liquid diffusion sources are painted on the wafer faces. The standard diffusion process is performed at 1255 °C and 24 h.

X-ray diffraction analysis was made after the standard diffusion process, for the both diffused wafer sides. For the P diffused wafer face, single peak rocking curves, having no significant differences as compared to the ones after grinding process, were obtained. In the same time, for the B-Al diffused face, two peaks rocking curves were obtained (figure 2a).The two peaks profile of the B-Al diffusion X-ray spectra stands for two different X-ray diffractions: on the substrate (the weaker peak ) and on the diffusion strained layer ( the larger peak ). A greater angular distance between

the two peaks is the result of a deeper strained layer.

The standard diffusion process was performed after different grinding processes (by 18μm and by 8μm $Al_2O_3$ grain size). No significant differences were observed on the rocking curves obtained on the both diffused wafer faces.

In order to obtain a diminishing of the diffusion induced strain, different diffusion processes were experimented. On the rocking curves of the B-Al diffused side, the distance between the two peaks and the width of the larger peak diminished with the reduction of the diffusion temperature and the growth of the process duration. The rocking curve obtained for 1210 $^0$C and 50 h diffusion process is presented in figure 2b.

Quantitative determinations were made. The parameters of gaussian spectra, which fit each of the both components of the obtained rocking curves, were determined. The elastic deformation of the lattice, normal to the surface, $m_\perp$ , and the internal residual strain, $\sigma$, were calculated for the both cases, "a" and "b", from figure 2. The results are the following:

- case "a" (standard diffusion process )

$$m_\perp = 8.40 \times 10^{-4}$$
$$\sigma = 10.96 \times 10^8 \text{ dyn/cm}^2$$

- case "b" (improved diffusion process )

$$m_\perp = 3.81 \times 10^{-4}$$
$$\sigma = 3.85 \times 10^8 \text{ dyn/cm}^2.$$

Therefore, the improved diffusion conditions led to an important reduction of the lattice deformation and of the internal residual strain.

## 5. Reliability improvement

RVT tests were performed for automotive diode batches processed at different grinding conditions but using the standard diffusion process. For the step between 18μm and 8μm grain size of $Al_2O_3$ abrasive, no significant differences of the test results were obtained.

Automotive diodes made with the improved diffusion process, performed at 1210$^0$ C / 50 h, were subjected at the RVT test, too. The failure percentages diminished by a factor between 4 and 8, as compared with the results for the standard diffusion process. The recorded failure percentages were below the 1% level. In operation, a similar reliability improvement was obtained.

## 6. Conclusions

The reliability of the automotive diodes is strongly dependent on the internal residual strain. The grinding process induces a high level strain but the obtainable reduction of this is negligible. The double diffusion process may induce an especially strong strain in the B-Al diffused side. A lower temperature / longer duration diffusion process led to an important reduction of the diffusion induced strain and to an important improvement of the automotive diodes reliability.

The X-ray diffraction spectra were an efficient method for pointing out of the internal strain and for monitoring its reduction.

## Acknowledgements

The authors are grateful to Mr. M. Udrea-Spenea for many helpful discussions.

## References

[1] Udrea–Spenea M. Automotive diodes reliability report. S.C. Baneasa S.A., Bucharest, Romania, 1996.

[2] Galateanu L., Stoica M. G., Bozdog C., Popa E. and Stoica A. D. Stress and strain in automotive diodes - a RVT, IR and XR study. Proceedings of International Semiconductor Conference (CAS), Sinaia, 1996, pp 439-442.

[3] Flewit P.E. J. and Wild R. K. In: IOP Publishing Ltd. Physical Methods for Materials Charaterisation. Philadelphia, 1994, pp 87-111.

PERGAMON

Microelectronics Reliability 38 (1998) 1335–1339

MICROELECTRONICS RELIABILITY

# Extrapolation of cosmic ray induced failures from test to field conditions for IGBT modules

C. Findeisen, E. Herr, M. Schenkel, R. Schlegel, H.R. Zeller

*ABB Semiconductors AG, Fabrikstrasse 3, CH-5600 Lenzburg, Switzerland*

**Abstract**

Recently it was found that biased high voltage IGBTs are more vulnerable to cosmic ray failures than devices like GTOs or thyristors. In order to obtain a reliable extrapolation from test to operating field conditions acceleration factors other than voltage are needed. For this purpose we exposed IGBTs and module diodes to an accelerator neutron beam and to the enhanced cosmic ray flux at 3580 m altitude above sea level. We conclude that neutrons are the most critical particles in cosmic rays with regard to potential device failures.
© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Data for cosmic ray induced failures of biased high voltage IGBTs are scarce compared to corresponding data for power devices like GTOs or thyristors. Collecting failure data from operating field conditions are impractical due to the necessary relatively long observation periods and are hampered by poorly defined parameters like shielding or uptime of the system. Accelerated tests can be performed with increased bias [1,2], with particle beams [3] or with increased cosmic ray fluxes. We exposed IGBTs and their corresponding freewheeling diodes (referred to as module diodes in the following) to an accelerator neutron beam and to the enhanced cosmic ray flux at high altitude and evaluated the results quantitatively.

## 2. Experimental Setup

We performed experiments with biased IGBTs and module diodes using the neutron beam 4FP30LA at LANSCE (Los Alamos Neutron Science Center) in Los Alamos, NM, USA. The neutron flux of this beam is eight orders of magnitude more intense than the neutron component of cosmic rays at sea level [4,5] (Fig. 1a). In addition both neutron energy distributions are similar up to 800 MeV. During the experiment the neutron beam passed through 12 or 24 planes of biased IGBT or module diode chips (Fig. 1b). The values of the neutron flux, the leakage current of all subassemblies (each consisting of four chips biased in parallel), and the bias monitor were logged for each run at a rate of 0.5 Hz. About 1000 IGBT and module diode chips of two different high voltage classes (2.5 and 3.3 kV) were tested at various biases. Neither the neutron flux nor the energy distribution was altered. The neutron flux reduction after 24 planes due to nuclear interactions between the neutrons and the subassemblies was less than 25% [6].

Commercial IGBT modules produced by ABB Semiconductors AG with the same chip technology as above were exposed to the enhanced

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00151-6

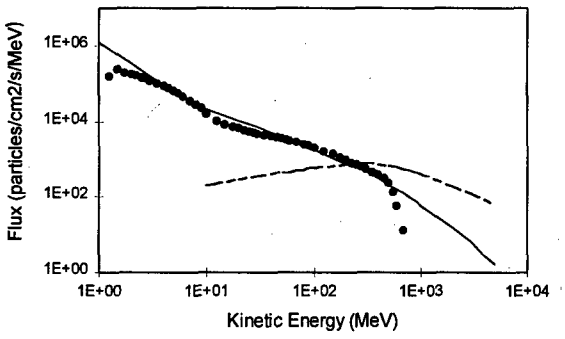



Fig. 1a: The neutron flux of the accelerator (•) compared to the neutron flux (—) and total muon flux (---) of cosmic rays at New York (sea level, 54°N geomagnetic latitude), both scaled with a factor $2x10^8$. Notice the missing neutron flux of the accelerator above the kinematic limit at 800 MeV. [4, 5].

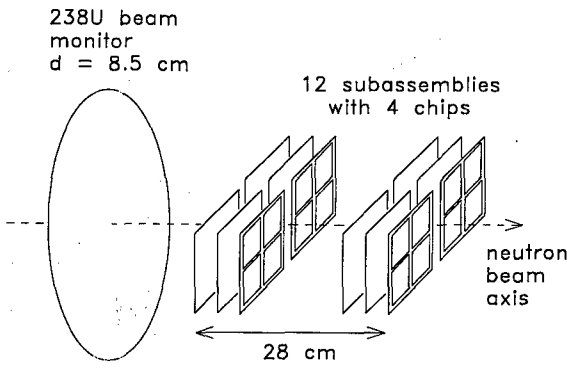



Fig. 1b: Experimental setup. The neutron beam of 8.5 cm in diameter and constant flux across the area passed through 12 or 24 planes of IGBT or diode chips. Four chips were mounted on a subassembly. There were two subassemblies per plane. A chip had an area of about 1 $cm^2$.

cosmic ray flux at 3580 m above sea level (Jungfraujoch, Switzerland). As a cross check the same types of modules were also exposed to cosmic rays at 400 m above sea level (Lenzburg, Switzerland), in order to calibrate the high altitude data. All tests were performed at room temperature.

## 3. Results

A failure was defined as one of the following two types of irreversible events. The first one consists of a sudden, irreversible increase in leakage current by more than a factor 10 (parametric failure), the second of the loss of blocking capability due to the formation of a filament of molten silicon (catastrophic failure). Nearly all of the observed failures were catastrophic.

The physical damage of the catastrophically failed chips was analyzed. If necessary the defect site was localized by liquid crystal hot spot detection. The typical failure pattern was molten silicon on the chip area (Fig. 2) and a melted channel extending all the way through the device.

Modeling of the acceleration law for particle beam or cosmic ray induced voltage breakdown is generally based on the assumption of a time-independent failure rate [7]. In order to verify this the observed lifetest data were statistically analyzed using cumulative Weibull plots (Fig. 3). The cumulative Weibull distribution

$$F(t) = 1 - \exp(-(t/\eta)^{\beta}) \quad (1)$$

is in this plot a straight line with shape parameter $\beta$ (slope) and the characteristic lifetime $\eta$. It is easily seen that $F(t=\eta) = 0.63$. The parameter t is either the number of neutrons passing through the chip before failure or the observed failure-free operating time.

For each run F(t) yields a shape parameter $\beta$ close to 1 (i.e. an exponential distribution law) (Fig. 3). Thus the assumption of the time-independent failure rate is consistent with the experimental observation.

All measured failure rates were related to the neutron flux at 54° N geomagnetic latitude and sea level (New York) [4]. The scaling factor used for the accelerator data was the ratio of the two neutron fluxes for $E_n > 10$ MeV: $6.8x10^7 = (3.80x10^5$ $n/cm^2/s) / (20$ $n/cm^2/h)$. The corresponding neutron flux of cosmic ray on the Jungfraujoch was about 240 $n/cm^2/h$.

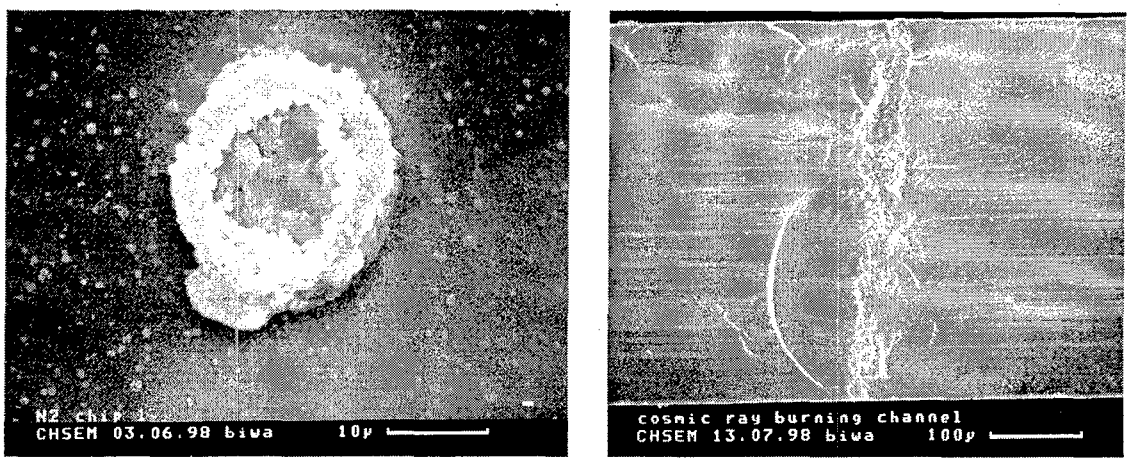



Fig. 2: **(left)** Neutron induced defect site. The melted channel is shown at the chip surface. The chip metallization layers were etched off. **(right)** Cosmic ray induced defect site. The melted channel is shown in a cross sectional view of the chip. This cross section was not polished.

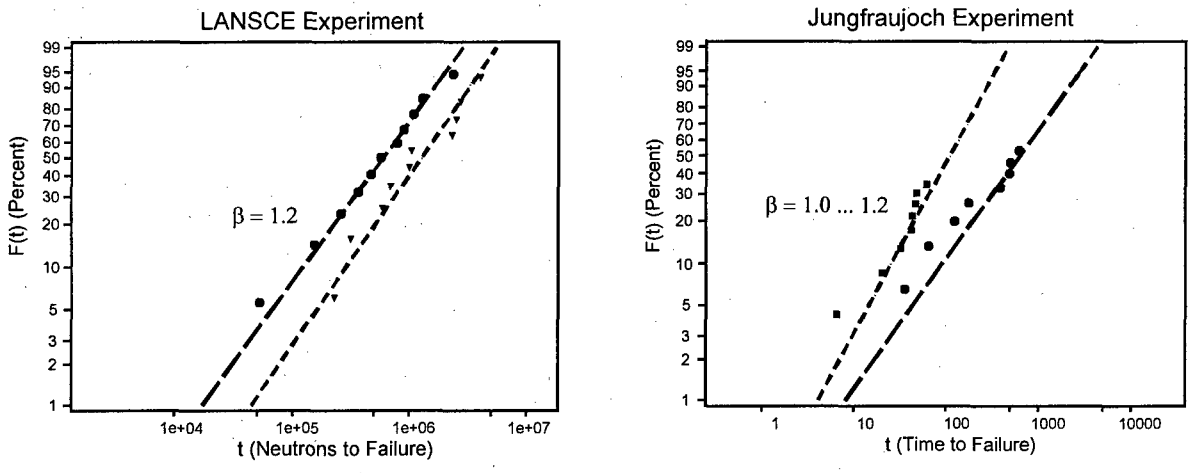



Fig. 3: Cumulative Weibull plot F(t) of failure times t for a population of IGBT substrates. t is either the number of neutrons passed until failure occurs or the time to failure. **(left)** Measurements using the accelerator neutron beam at two different biases. **(right)** Measurements on the Jungfraujoch at two different biases.

## 4. Discussion

The scaled experimental failure rates were normalized to the chip area and n-base resistivity and plotted as function of the parameter S which is proportional to the maximum electrical field in the device (Fig. 4). If the local failure probability is only a function of the local electrical field and if the field is linearly dependent on position in the n-base space charge zone, then the plot of Fig. 4 should result in a universal curve [2, 7]. The parameter S is for a non punch-through device given by $S=\sqrt{(V/\rho)}$, where V is the bias in volts and $\rho$ is the n-base resistivity in $\Omega$cm. For a punch-through device $S = 0.2736V/t + 0.8972t/\rho$ where t in $\mu$m is the n-base thickness.

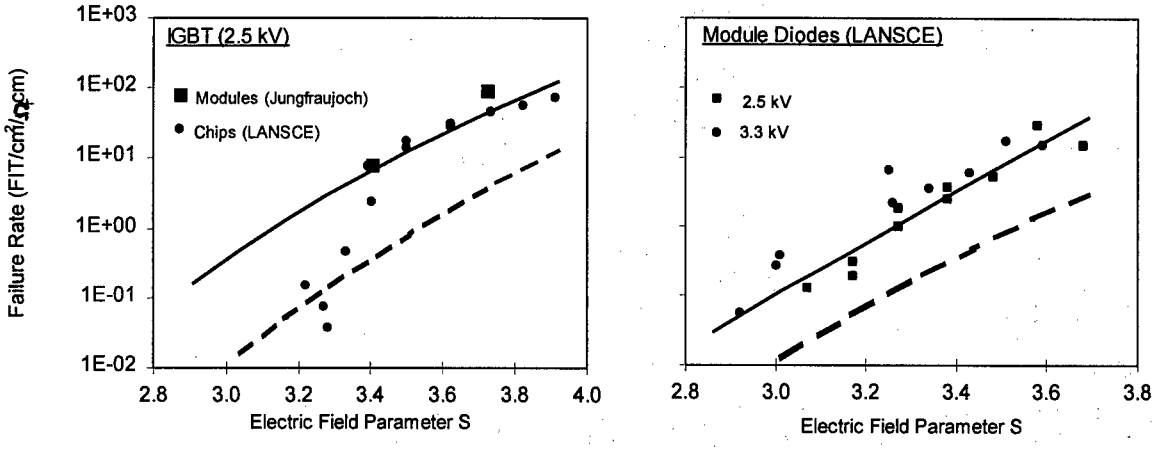



Fig. 4: Measured failure rates for IGBT and module diodes as a function of the electric field parameter S and normalized to the chip area and the n-base resistivity ρ. The failure rates measured at LANSCE and on the Jungfraujoch are scaled to sea level neutron flux. The dashed line 'master curve' is the result of a model for devices with deep junctions like GTO, GTO diodes or thyristors. **(left)** The failure rate for IGBTs as measured at LANSCE and on the Jungfraujoch. The solid line is the modified 'master curve' which is corrected for the difference in junction profile between GTO and IGBT. **(right)** The failure rate for module diodes as measured at LANSCE. The solid line is the modified 'master curve' which is corrected for the difference in junction profile between deep diffused diodes and module diodes. 1 FIT corresponds to 1 failure every $10^9$ device•hour.

Numerical studies show that the deviation from a triangular or trapezoidal field distribution in the vicinity of a deeply diffused junction leads to a parallel shift of about an order of magnitude for voltages of 2 - 4.5 kV. Earlier experiments with thyristors, GTO's and wafer scale diodes gave failure rates which did fall on a master curve [2, 7], but the "true" master curve for devices with an abrupt junction will be shifted upwards.

The diodes of Fig. 4 have a nearly abrupt junction, their failure rates follow a curve parallel to the GTO curve and the shift compared to the GTO curve is consistent with the expected difference between an abrupt and a deep junction.

The IGBT results are more complex. The IGBT has also a nearly abrupt junction but furthermore the electrical field distribution is not one-dimensional as in bipolar power devices such as diodes, GTO's and thyristors. It has been argued [7] that 2-D effects in the field distribution do not significantly affect the failure rate. A similar behavior as for diodes would then be expected. This is consistent with the failure rate observed at high S values.

At smaller S values (S < 3.4) the failure rate has a cutoff. Similar cutoff effects have been reported also for diodes [3]. The cutoff cannot be explained by a model in which the total failure probability is an integral over local probabilities which depend only on local field [2, 7]. Such a model would predict the cutoff to occur at a constant S, which is not observed. More work is required to understand the physics of the cutoff.

It had been suggested in [3] and later in [7] that cosmic neutrons are the prime causative agens for cosmic ray induced failures. In our work we demonstrate a quantitative agreement between accelerator data and data from high altitude experiments.

We strongly believe that the cutoff observed in accelerator experiments is real and would also be observed in devices exposed to cosmic rays. If the cutoff would be due to the kinematic energy cutoff of accelerator neutrons at $E_n$ > 800 MeV or the missing muons with E > 1000 MeV [4], then it would be observed at the same S value for all devices. This is not the case.

## 5. Conclusions

Models which are based on a failure rate controlled by either local nucleation of an avalanche [2, 7] or a scaling with the ionization integral in the n-base [1, 3] can describe the overall failure behavior rather well, but not the cutoff.

Our finding, that the Los Alamos results can be quanitatively related to cosmic ray failures give credibility to the extrapolation procedures used in [1, 2, 3, 7] to predict cosmic ray induced failures in the field. The extrapolations give upper limits for the failure rates because they do not include a cutoff. Until the physics of the cutoff is better understood, it is a safe approach to ignore the cutoff in the extrapolation unless it has been experimentally verified for the actual device under test.

## Acknowledgments

Parts of this work were supported by the Brite-Euram program RAPSDRA and by the Swiss Federal Office of Education and Science. Prof. H. Debrunner (Swiss Alpine Research Institute Jungfraujoch and University of Berne) generously provided access to the Jungfraujoch facilities. This work has benefited from the use of the Los Alamos Neutron Science Center at the Los Alamos National Laboratory. This facility is funded by the US Department of Energy under Contract W-7405-ENG-36.

## References

[1] H. Kabza et al., Proc IEEE Intern. Symp. Power Semicond. Devices and ICs, Davos, pp. 9-12, 1994

[2] H.R. Zeller, Proc IEEE Intern. Symp.Power Semicond. Devices and ICs, Davos, pp. 339-340, 1994

[3] P. Voss et al., Proc IEEE Intern. Symp. Power Semicond. Devices and ICs, Weimar, pp. 169-172, 1997

[4] IBM Journal of Research and Development, Vol. 40, No. 1, pp. 19-39, January 1996

[5] S. Wender, LANSCE, private communication

[6] J.F.Ziegler and W.A.Lanford, J.Appl.Phys. **52**, No 6, pp. 4305-4312, 1981

[7] H.R. Zeller, Microelectron. Reliab., Vol. 37, No. 10/11, pp. 1711-1718, 1997

PERGAMON

Microelectronics Reliability 38 (1998) 1341–1345

MICROELECTRONICS RELIABILITY

# Thermomechanical deformation imaging of power devices by Electronic Speckle Pattern Interferometry (ESPI)

K. NASSIM*, L. JOANNES, A. CORNET,
Laboratoire FYAM, Université de Louvain, 1348 Louvain-La-Neuve, Belgium
* On leave from University Chouaib, El Jadina, Marocco

S. DILHAIRE, E. SCHAUB and W. CLAEYS
Laboratoire de Caracterisation de composants Electroniques, CPMOH and IXL
Université Bordeaux 1 33405 Talence Cedex, France
Fax : 33 556 846 970, Email : wclaeys@frbdx11.cribx1.u-bordeaux.fr

**Abstract**

We present an original imaging method to measure the three components of the surface displacement of working power devices with a nanometric resolution. The method takes advantage of the speckle structure of the analysed object recorded on a CCD camera. This method is complementary of IR thermography and provides interesting information concerning stress and reliability in power devices.

© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

The reliability of power devices depends strongly on their working temperature and the related thermal dilatation. Power devices are assemblies made of the superposition of different layers with different dilatation coefficients. The mismatch of the dilatation coefficients gives rise to stress into the layers. This stress induced by the functioning of the device results in the deformation of the structure.
Many tools have been developed to get information about the thermal and the thermomechanical behaviour of power devices. The most used is IR thermography which allows to measure the surface temperature of the component. Thermomechanical effects can be studied by laser interferometry [1, 2] as it allows to measure the normal surface displacement of a point of the device. Surface normal displacement imaging is possible by scanning the device but it is time consuming.

In this paper we propose an optical imaging method based on speckle interferometry [3]. Such techniques based on upon the analysis of laser granularity (speckle) are today very promising as whole field optical methods for surface deformation testing. High accuracy measurement of the three components of the deformation vector of an optical rough surface can be obtained directly from a set of images. Coherent light imaging, CCD cameras, PC image digitising boards and image processing software are the only technical ingredients necessary to carry out such methods. These techniques are widely spread for the study of macroscopic objects, like airplanes, but have not yet been used to study electronic devices.
Power devices, satellite and car electronics for example undergo severe thermal solicitations, the understanding of thermomechanical consequences is crucial in order to meet reliability issues.
In this paper we propose this new method of thermomechanical analysis of electronic devices and apply it to determine the out-of-plane (z) and in-plane (x) components of displacement of each point of the surface (2x3 $mm^2$) of an active MOS transistor in thermal dissipative equilibrium.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00079-1

Absolute values of out-of-plane displacement are obtained with a high accuracy technique based on interferometric measurement of the displacement of a focused laser beam [1].

## 2. Principle and measurement techniques

Two different optical set-ups are used to measure both out-of-plane and in-plane surface displacement. They are presented respectively in figures 1 and figure 2. The purpose is to measure with a simple procedure the displacement vector of each point of the surface of a device due to its running in the nanometer range.

Speckle techniques are based on limited bandpass coherent imaging of rough surfaces. The limited resolution leads to the granular aspect of objects. The size of the grains of light is related to the numerical aperture of the imaging system which is controlled by the diaphragm in front of the lens of the CCD camera. The diaphragm is adjusted to obtain a spatial sampling frequency three time greater than the highest speckle pattern frequency. The spatial sampling frequency is determined by the size of the CCD camera pixels. Under these conditions small dots of light (speckle) are detected by the CCD camera while imaging the device under test illuminated by laser light. These dots are emitted from corresponding small areas on the object.

### *2.1 Out-of-plane measurement*

Out-of-plane displacement measurement is done by detecting the phase shift undergone by the light coming from each dot of the moving surface. The phase of the speckle pattern is coded by the superposition of a reference wave front coming from the reference arm (see figure 1). This superposition of the diffused light from the object and the reference beam leads to a new speckle pattern distribution on the CCD camera due to the interferences of the two beams. We can then consider the rough surface made of small mirrors whose size is determined by the speckle grain size. The displacement of grains out of their plane is measured by interferometry on each dot, this means comparing the light amplitude to the interference pattern. General procedure for data acquisition and process is made in four steps.

The phase of each speckle dot relatively to the reference wave is first calculated. Therefore a set of five intensity images, with controlled phase shifts of the reference field are recorded, they produce the interference pattern of the dot under study and allow to extract the speckle dot phase for the non running device.

When the transistor is operated and at steady state, a new set of five images is taken in order to calculate the new phase of the speckle dots relatively to the reference beam. Phase variation due to out-of-plane displacement is determined by subtraction of the reference phase image. This is done for a first speckle dot and contains a $2\pi$ phase indetermination. An unwrapping procedure made by the comparison of neighbouring dots allows to remove subsequential $2\pi$ phase indetermination. This means that the relative position change of all the dots of the surface are determined in an absolute way with respect to the reference dot. This provides a set of calibrated metric displacement values.

Results show the spatial variation of displacements. By measuring the thermal expansion using a laser interferometer [1] on one dot, one can determine the total displacement of the reference dot and hence the complete absolute displacement of all the points by using the speckle displacement map.

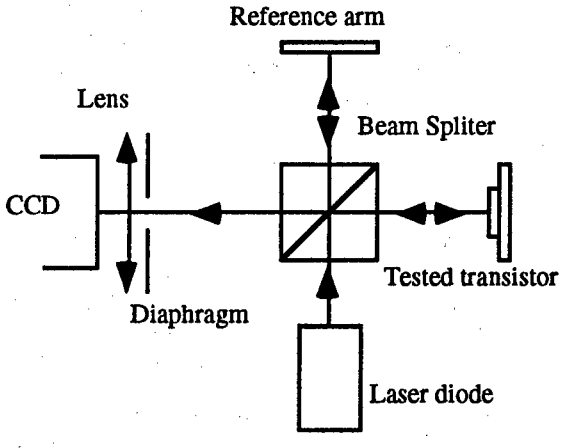



Figure 1 : Out-of-plane configuration

### *2.2 In-plane measurements*

In plane measurements are done by illuminating the object a sinusoidal light amplitude pattern. This is obtained by shining two plane wave laser beams, which interfere upon the sample under test and produce a sine wave amplitude pattern. The step of the illumination grating depends on the light

wavelength and the angle α between the two beams (see figure 2). The CCD camera records the speckle pattern encoded in a light amplitude grating. The displacement is obtained by determining the position of the grains inside the fixed grating. The procedure for data acquisition is quite similar to the one of out-of-plane measurements as here too five recordings are made. The phase shifts are obtained by small changes of the wavelength of a tunable laser diode in the unbalanced Michelson interferometer (see figure 1). Two sets of five measurements, one for the non running device and one for the running provide the displacement of the analysed dot. The same $2\pi$ phase indetermination for the reference dot occurs and an unwrapping procedure is used. The small amplitude of the measured displacements together with tests at increasing power levels allow us to obtain absolute displacements.

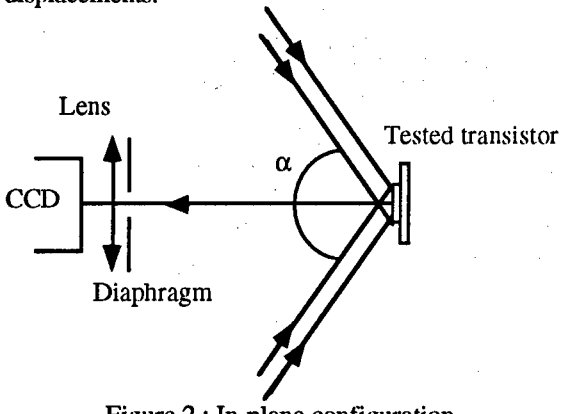



Figure 2 : In-plane configuration

## 3. Results

A MOS power transistor has been studied. Figure 3 shows the top view of the die. The top of the package has been removed revealing a $2\times3\mu m^2$ chip. The left electrode is the gate, the right one is the source, the drain is the back side. When the transistor is switched on, it is supplied by a constant controlled current. During operation the transistor dissipates 1 watt. Figures 4 and 5 show the surface displacement respectively in-plane along the x-axis and out-of-plane along the z-axis produced under steady state operation.

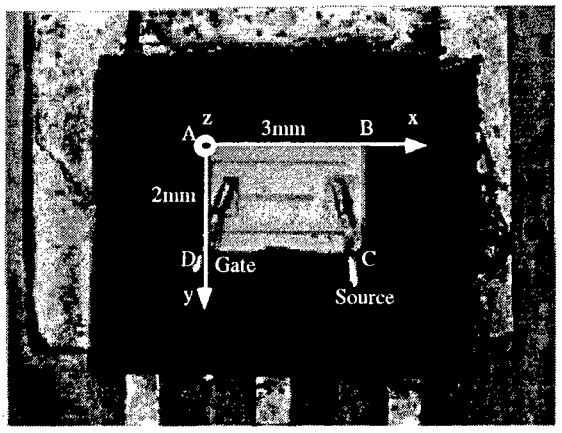



Figure 3 : Top view of the Power MOS transistor

The in-plane surface displacement shows a global movement in the positive x-direction (see figure 4). The amplitude in figure 4 shows the displacement of a measured point (x,y) in the x direction during operation with respect to its initial position before operation. This representation is valid provided the displacements are small compared to the dot sizes analysed. The further the point is from the y-axis the greater the amplitude of the displacement. This is typical of surface expansion where the AD border is almost fixed with respect to the analysing system and all the points then move away from this line.

For a better understanding of this type of surface displacement representation, the following example can help. The free expansion of a metal foil would produce a x-displacement profile showing as an inclined plane in the 3D representation of figure 4. This is mainly the case for the transistor with some stress showing up.

The maximum amplitude movement is 60nm along the BC direction while it is 25nm along the AD direction. Similar measurements can be performed for the in-plane displacement in the y direction.

Figure 5 shows the out-of-plane z-axis displacement, a bending of the structure together with a global tilt of the surface of the transistor is observed.

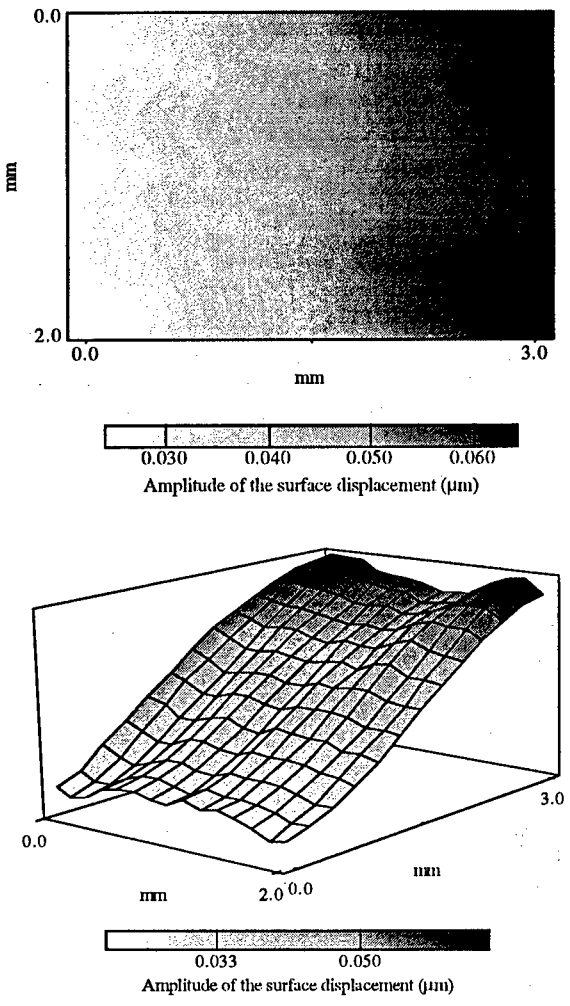



Figure 4 : In-plane x-axis displacement cartography

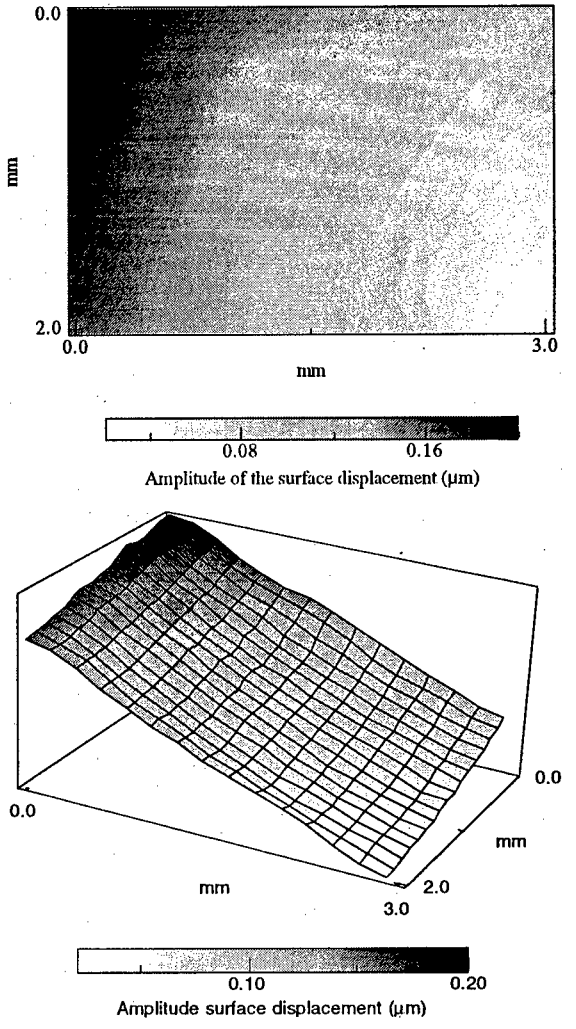



Figure 5 : Out-of-plane z-axis displacement cartography

The point A (see figure 3) practically doesn't move while the amplitude of the movement of the point C is 0.2µm. The movement of one point of the surface integrates the contribution of all the points on the vertical line under the surface and also the contribution of the stress induced by the functioning of the device.

The electrodes are not visible on both images. It means that the surface displacement of the chip is led by the diffusion of heat in the layers under the die. The electrodes dilatation dominates during the first milliseconds of the diffusion process, when the heat flux does not yet cross the first interface. A few seconds after the start of the excitation, the thermomechanical behaviour is dominated by the dilatation of the different layers of the transistor.

The first instants of the dilatation cannot be resolved by the CCD matrix (25Hz) but it is possible to record the surface displacement out-of-plane of one point of the MOS transistor using our high sensitive stabilised homodyne interferometer [1]. Figure 6 shows the displacement due to thermal expansion of the component. The transient surface displacement is recorded up to 20 seconds. This information is used to determine the total absolute displacement value of the reference point. The absolute displacement of the reference point can be obtained by directly from the CCD data for slow varying surface displacements. In the case of the

transistor this means to move to the steady state by slow power increase instead of a step function. Absolute values obtained from the interferometer agree with those obtained by ESPI.

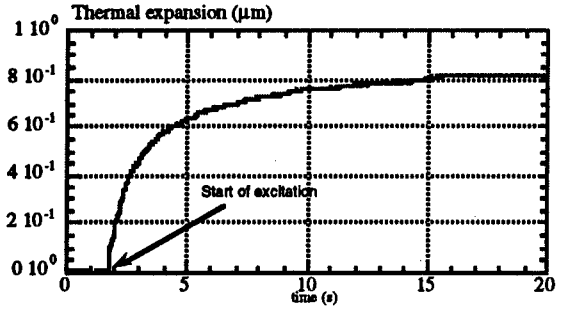



Figure 6 : Thermal expansion obtained at one corner of the power transistor by laser interferometry

ESPI displacement resolution is typically better than 50nm for both in-plane and out-of-plane measurements and is obtained with the use of optimised phase calculation algorithms on speckled data.

## 4. Conclusion

We have presented an original and powerful tool to image and determine the in-plane and out-of-plane deformation of working power devices. The method takes advantage of the speckle structure seen with a laser light-CCD camera imaging system of a power device. Speckle interferometry allows to measure the displacement of each speckle dot for both in-plane and out-of-plane movement. The knowledge of the displacement vector of each points of a surface with this imaging technique provides interesting information to understand deformation, stress and failure mechanisms in power devices. The use of such a technique in reliability will help a lot to validate thermomechanical simulations.

## References

[1] W. Claeys, S. Dilhaire and V. Quintard.
Laser probing of thermal behaviour of electronic components and its application in quality and reliability testing.
Microelectronic Engineering, 24, p. 411-420, 1994.

[2] W. Claeys, S. Dilhaire, and E. Schaub.
Laser probing techniques and methods for the thermal characterization of microelectronic components.
Thermal Management of Electronic Systems, J.-P. Bardon, E. Beyne, J.-B. Saulnier Editors. ELSEVIER Paris, pp227-237, 1998. ISBN2-84299-033-1.

[3] C. Wykes
Use of Electronic Speckle Pattern Interferometry (ESPI) in the measurent of static and dynamic surface displacement.
Optical Ingineering, Vol. 21 (3), pp 400-406, 1982.

PERGAMON

Microelectronics Reliability 38 (1998) 1347–1352

MICROELECTRONICS RELIABILITY

# On the effect of power cycling stress on IGBT modules

P. Cova[a], F. Fantini[b]

[a] *University of Parma, Dipartimento di Ingegneria dell'Informazione, Viale delle Scienze, 43100 Parma, Italy*

[b] *University of Modena, Dipartimento di Scienze dell'Ingegneria and INFM, via G. Campi, 213/b, 41100 Modena, Italy*

**Abstract**

IGBT reliability is becoming of great relevance, due to the range of application of these devices. Nevertheless, no standard test methods have been established, in order to evaluate their power cycling reliability. On this paper we report on the effect of $\Delta T$ and $T_{jmax}$ on the power cycling capability of IGBT dice, by means of a matrix of stress cycles with different values of $\Delta T$ and $T_{jmax}$. Failure analysis has been performed, in order to understand the failure mechanisms induced by the stress.
© 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

During the last years the importance of IGBT reliability has continuously increased, due to the widespread use of these devices in many fields, including lighting, power conversion, motor drivers and electronic ovens. To asses their reliability in these applications, power cycling is the most suitable stress test, because the devices are operated in conditions similar to those encountered in the field. In effect, the main cause of failure is the repetitive thermal cycling, which occurs when power cycling is applied [1]. Failure modes of IGBT modules were identified in the past as to be related to Al wires debonding, due to Al reconstruction [2], to thermo-mechanical fatigue produced by the repetitive expansion cycles of the wires [3] and to thermo-mechanical problems in the die attach [4, 5]. Recent papers have reported substantial reliability improvements, thanks to new technological solutions [3, 6].

In spite of the fact that reliability is a key factor for the development of the IGBT technology, only few data are available and no standard test methods have been defined to evaluate the power cycling reliability and the effective impact of the various stress paramenter is not yet established [3, 6, 7]. On this paper we report on the effect of $\Delta T$ and $T_{jmax}$ on the stress level during thermal cycling of IGBT modules, by means of the study performed on a matrix of stress cycles employing different values of $\Delta T$ and $T_{jmax}$. Failure analysis is also used, to understand the failure mechanisms induced by the different stresses.

In paragraph 2 the measurement set-up is described, paragraph 3 reports on stress cycles results and paragraph 4 deals with the failure analysis. Discussion and conclusions follow.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00081-X

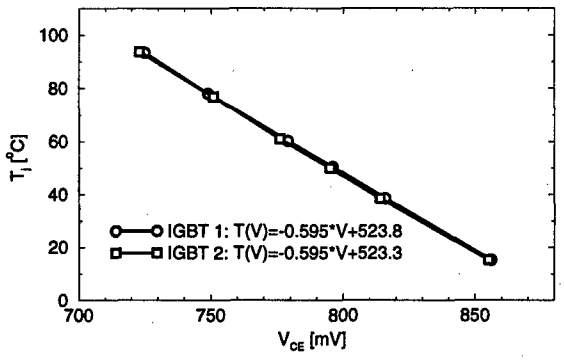



Fig. 1. Calibration curve ($T_j$ versus $V_{CE}$ at $I_C = 100$ mA) for two of the tested IGBTs.

## 2. Measurement set-up

The tested samples, purchased from various manufacturers, were commercially available medium power devices, packaged in modules containing 6 IGBTs in a three-phase inverter configuration, with fast free-wheeling diodes. Every IGBT was guaranteed for a 10 A, 600 V operation. We report here on the power cycling tests performed on two different lots of samples from the same manufacturer, named in the follow A (from an older production) and B (the newer one).

To submit the modules to power cycles, we designed and built-up a system able to supply a predefined square-wave collector current into a single IGBT and to measure at the same time its junction temperature. Fig. 2 shows the block diagram of the measurement equipment. We selected the collector-emitter saturation voltage as a suitable TSP (Temperature Sensitive Parameter), because it exhibits a linear dependence with respect to the junction temperature at a given measuring current [8, 9, 10]. To determine the relationship between $V_{CE}$ and T in the range from 20 °C up to 125 °C (calibration curve, see Fig. 1), we injected the measuring current $I_m$=100 mA in the collector and recorded the corresponding $V_{CE}$ value. The measurements were performed under both continuous and pulsed currents and no differences were observed, so that we can conclude that this level of measuring current does not appreciably influence the junction temperature. The TSP is monitored during the fast electric transient between the on-off conditions, before the beginning of the slow thermal transient, when the current level is equal to the measuring current (see Fig. 3). For these measurements, a current power supply, which is able to supply a current up to 20 A with an arbitrary waveform, in accordance with a set-point signal, was built. All measurements are controlled by means of a PC driven board.

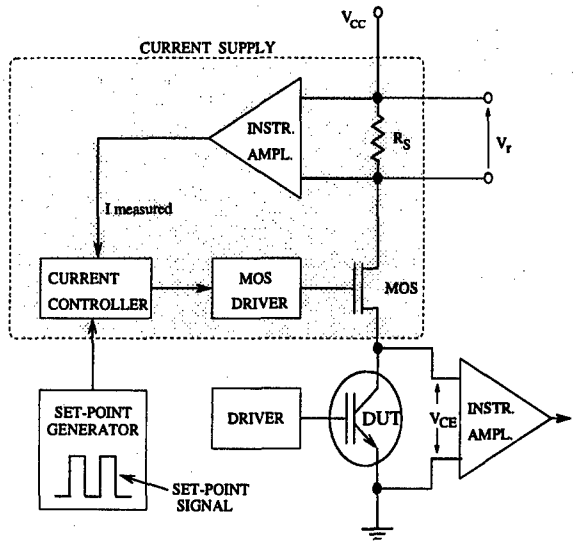



Fig. 2. Block diagram of the circuit used to submit the samples to power cycling and measure the junction temperature.

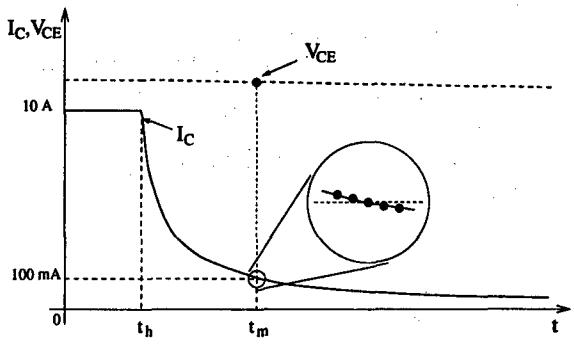



Fig. 3. Magnification of the collector current waveform, showing the $V_{CE}$ sampling when $I_C = 100$ mA.

During the test, the samples are stressed by a rectangular wave collector current and, at every period, the temperature is monitored. The gate voltage is kept high (12 V) during the operation. The module is placed in a thermally controlled chamber, able to establish the ambient tempera-

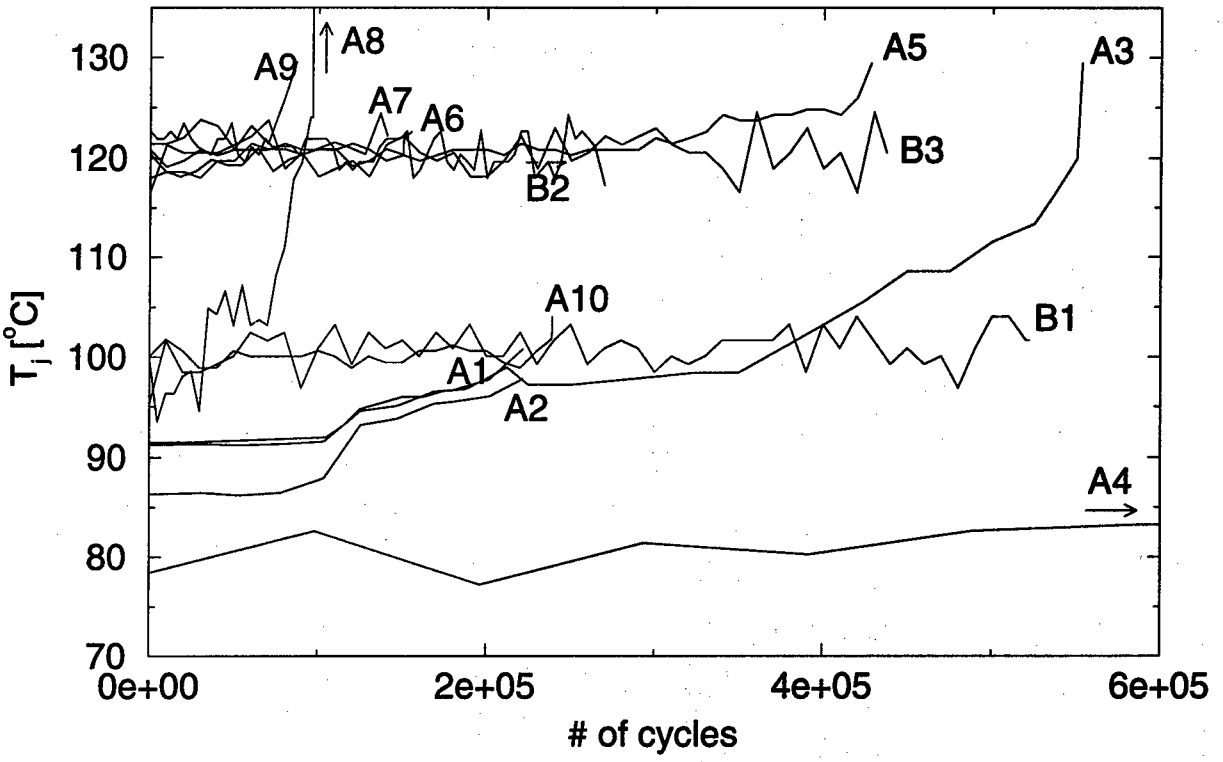



Fig. 4. Junction temperature versus the number of cycles for the tested samples.

ture, by forced air ventilation, from 0 to 150 °C. The temperature of the device depends on the duty-cycle of the square waveform. We set the high current level, $I_h$, equal to the maximum continuous current allowed for the devices ($I_h$ =10 A), the low current level above the measurement level (for instance 50 mA) and, by a trigger system, we measure $V_{CE}$ when $I_C = I_m$ =100 mA.

If, as in our case, the sampling time (of the order of 10 $\mu$s) is much shorter than the time constant of the current supply (hundreds of $\mu$s), the error we made in the $T_j$ measurement (i.e. the difference between 100 mA and the real value of the current during the $V_{CE}$ measurement) is negligible. Furthermore, as illustrated by Fig. 3, the time between the heating current switch time ($t_h$) and the $V_{CE}$ measurement time ($t_m$) is of the order of hundreds of $\mu$s, much shorter than the thermal time constant of the device (hundreds of ms or more).

The test on lot A samples was performed by keeping constant the duty-cycle of the collector current: an increase of the junction temperature was generally observed in the final part of the test, as shown in Fig. 4. This behaviour was due to the increase of the thermal resistance, probably because of the degradation of the interfaces between the different materials employed. To avoid this positive feedback, which could affect the life-tests results, we introduced a duty-cycle automatic software control, able to keep constant the junction temperature of the device during the whole test. The start-up of this correction and its amount are monitors of the device degradation and we recorded, together with the values of $V_{CE}$, the duty-cycle variations during the stress.

## 3. Stress cycles

Stress tests on the modules have been performed by subjecting the devices to different power cycles, all with $I_m$ =10 A, but with different duty-cycle, which sets the thermal excursion. Table 1 summarizes the parameters of the stresses performed. For samples from lot B we

Table 1. Main parameters of the stresses performed.

| Sample | $\Delta T$ (°C) | $T_{jmax}$(°C) | Period (s) | Duty-cycle (% ) | # of cycles | Failure |
|---|---|---|---|---|---|---|
| A1 | 65 | 90 | 2.0 | 50 | 221,000 | no |
| A2 | 65 | 90 | 2.0 | 50 | 221,000 | no |
| A3 | 65 | 90 | 2.5 | 12 | 553,000 | emitter bond lift-off |
| A4 | 50 | 82 | 3.0 | 80 | 3,300,000 | emitter bond damage |
| A5 | 50 | 120 | 3.5 | 9.4 | 428,500 | emitter bond damage |
| A6 | 60 | 120 | 3.5 | 35 | 155,000 | emitter bond damage |
| A7 | 60 | 120 | 3.5 | 35 | 140,700 | emitter bond damage |
| A8 | 70 | 100 | 21.2 | 94 | 96,300 | emitter bond lift-off |
| A9 | 65 | 120 | 5.7 | 33 | 86,700 | emitter bond lift-off |
| A10 | 55 | 100 | 3.0 | 26 | 239,000 | emitter bond damage |
| B1 | 60 | 100 | 3.5 | 8 | 522,400 | to be analyzed |
| B2 | 70 | 120 | 3.8 | 22 | 270,100 | to be analyzed |
| B3 | 60 | 120 | 3.5 | 12 | 437,700 | to be analyzed |

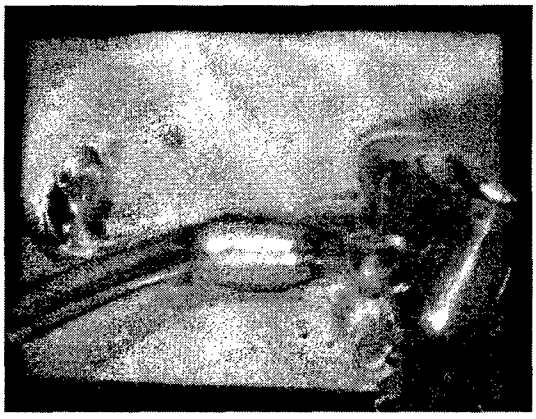


Fig. 5. Photograph of the failed sample A3, showing the break in the emitter bond.

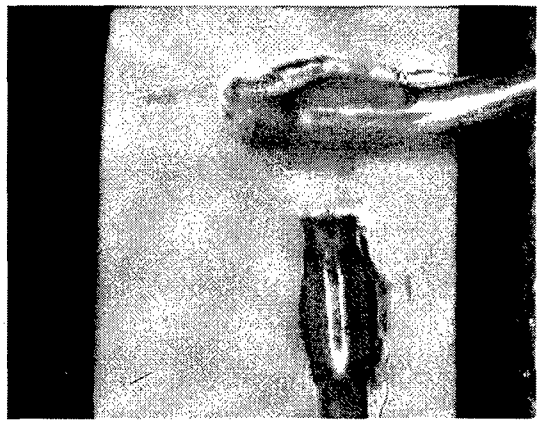


Fig. 6. Photograph of the failed sample A4; the damage in the emitter bond is less evident.

have reported the duty-cycle at the beginning of the stress.

Fig. 4 reports on the values of $T_j$ monitored during the stress for all our samples. For most of the samples from lot A both the minimum and maximum temperatures increased in the last part of the test, probably due to the degradation of the die attach thermal resistance, but the thermal excursion was maintained constant.

For two samples (A1 and A2) we stopped the stress after the observation of an increase of $T_{jmax}$ of about 10% and we acquired again the calibration curve, finding a shift of about $5 \div 6$ °C. It means that a corresponding variation in the device temperature may occur during the test, in spite of the duty-cycle correction.

## 4. Failure analysis

The failed samples were submitted to electrical tests, in order to verify the presence of short or open circuits. All the samples showed an open circuit between emitter and collector. The physical failure analysis was performed by means of optical and SEM failure analysis, after opening

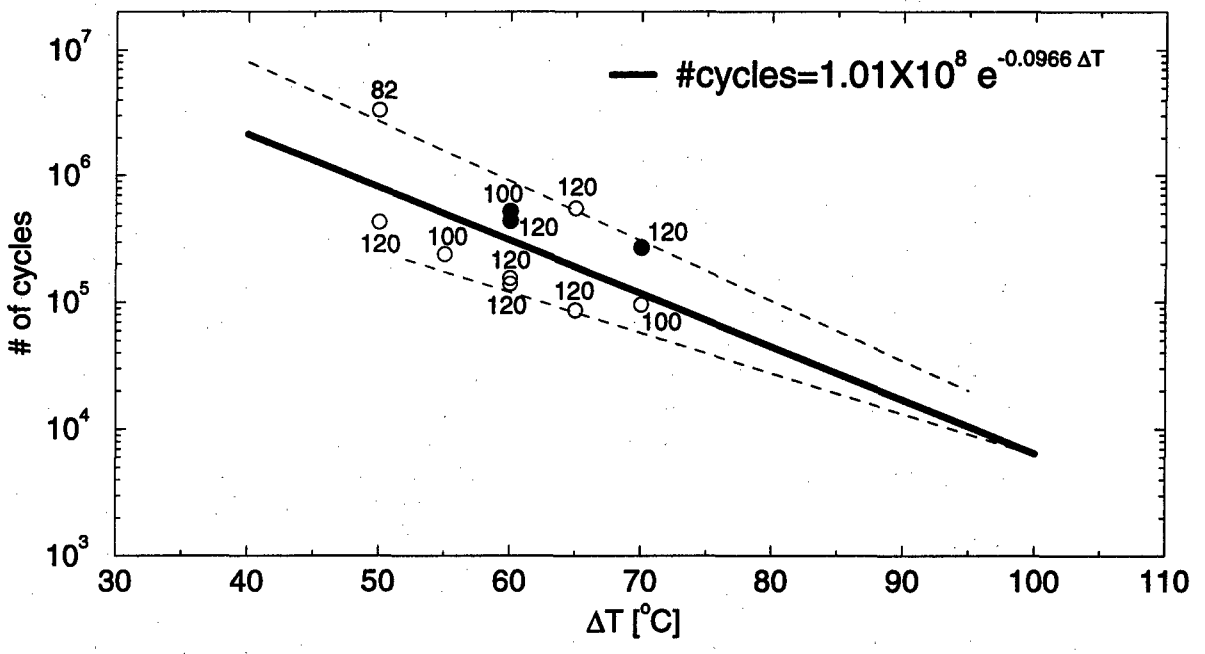



Fig. 7. IGBT lifetime versus $\Delta T$ for the tested samples. The numbers close to the symbols are $T_{jmax}$. Dashed lines limit the range of results from many groups, while the bold line represents the interpolation of the whole set of data.

the package and removing the insulating gel. For all the devices we observed a break or a damage of the emitter bond. Fig. 5 and Fig. 6 show two different levels of device degradation: sample A3 shows a large shift of the bonding wire; in sample A4, on the contrary, the vertical displacement of the wire is less evident, because the photograph was taken at an early phase of degradation.

Optical and SEM inspection did not show any macroscopical damage on the die attach, but microscopical internal degradation of the interfaces cannot be excluded at this stage of the analysis.

## 5. Discussion

The measurements on devices from lot B were performed after updating the system with the automatic duty-cycle correction and this improvement has to be taken into account when we compare these results with those obtained on the lot A. To this purpose we have to look at the behaviour of the junction temperature of the samples from lot A during the stress, from which the onset of thermal resistance increase appears.

More data have to be collected, in order to perform a statistical analysis, but from these preliminary results some conclusions can be drawn.

### 5.1. Dependance of the lifetime versus $\Delta T$

In Fig. 7 we show a collection of lifetime results versus $\Delta T$ from our tests and data from literature [3, 6, 7]. The results of life-tests on our samples are represented by the circles (open for lot A, filled for lot B) while the dashed lines show the limits for the data collected from the literature. The bold line is a tentative interpolation of all these data. The sample lifetime seems to be exponentially related to $\Delta T$ and we can extrapolate a device lifetime longer than $10^6$ cycles for the typical application range ($\Delta T = 20 \div 40\ ^\circ C$).

### 5.2. Influence of $T_{jmax}$

The values of $T_{jmax}$ for our tests are reported close to the circles in Fig. 7. From the analysis of the whole set of data it seems that a high value of $T_{jmax}$ ($> 120\ ^\circ C$) drastically decreases the

lifetime, while at lower $T_{jmax}$ a clear dependence is not exhibited.

### 5.3. Failure modes

Even though electrical and optical analysis of the devices demonstrated that the emitter bond damage was the main failure mechanism, the increase of the junction temperature and the shift of the calibration curve of the samples from lot A, clearly indicated that the degradation of the thermal resistance between the die and the heat-sink (which is composed by various layers of different materials), is not negligible. This phenomenon appears mostly related to $\Delta T$, while no particular correlation was found with $T_{jmax}$.

### 5.5. Technological improvements

Although this observation is at present based on few data, our preliminary measurements on samples from lot B confirmed the influence of the technological improvements on the IGBT modules lifetimes. A correlation with particular technological aspects cannot be made at present because constructive details are not known.

## 6. Conclusions

Power cycling is the most largely employed test for the reliability evaluation of power IGBTs. The thermal excursion $\Delta T$ has been found to be the more important paramenter in this type of stress, and may be the cause of damage of the soldering layers and bond wires. From the analysis of our data and the results from other groups we can assert that the power cycling lifetime is exponentially related to $\Delta T$, so that it is possible to evaluate the field lifetime, by means of accelerated test, if the thermal excursion during application is known. For normal application, where thermal excursions range from 20 to 40 $°C$, IGBT lifetimes longer than $10^6$ cycles can be extrapolated.

The maximum temperature operation $T_{jmax}$ has a lower impact on the lifetimes, if it does not reach temperatures > 120 °C. This situation limits the possibility of performing accelerated tests at very high $T_{jmax}$ without inducing failure modes, which are specific of the test conditions.

We have observed the emitter wire bonding lift-off or damage as the main failure mechanisms, although the degradation of the die attach is probably the root cause of the failure mechanism.

## References

[1] Ohga K. Failure analysis of bonding wires in power transistor modules. Proc. ISTFA, 237-247, Los Angles, USA, 1991.

[2] Malberti P, Ciappa M, Cattomio R. A power-cycling-induced failure mechanism of IGBT multichip modules. Proc. ISTFA, 163-168, Santa Clara, USA, 1995.

[3] Sankaran V A, Chen C, Avant C S, Xu X. Power cycling reliability of IGBT power modules. IEEE Ind. Appl. Soc. Annual Meeting, New Orleans, USA, 1997.

[4] Franceschini G, Fantini F, Salini D, Bricca G, Muschitiello M. Reliability evaluation of power SCR's. Proc. ESREF, 51-54, Scwäbisch Gmünd, Germany, 1992.

[5] de Lambilly H, Keser H O. Failure analysis of power modules: a look at the packaging and reliability of large IGBTs. IEEE Trans. on CHMT, 16 (1993) 412-417.

[6] Sommer K, Göttert J, Lefranc G, Spanke R. Multichip high power IGBT-modules for traction and industrial application. Proc. EPE, 1.112-1.116, Trondheim, Norway, 1997.

[7] Hamidi A, Coquery G. Effect of current density and chip temperature distribution on lifetime of high power IGBT modules in traction working conditions. Microelectron. Reliab. 37 (1997) 1755-1758.

[8] Baliga B J. Evolution of MOS-bipolar power semiconductor technology. Proc. of IEEE, 76 (1995) 409-418.

[9] Baliga B J. Temperature behavior of insulated gate transistor characteristics. Solid State Elect. 28 (1985) 289-297.

[10] Cova P, Ciappa M, Franceschini G, Malberti P, Fantini F. Thermal characterization of IGBT power modules. Microelectron. Reliab. 37 (1997) 1731-1734.

PERGAMON

Microelectronics Reliability 38 (1998) 1353–1359

MICROELECTRONICS RELIABILITY

# Temperature measurements and thermal modeling of high power IGBT multichip modules for reliability investigations in traction applications

A. Hamidi[a,c], G. Coquery[a], R. Lallemand[a], P. Vales[b], J.M. Dorkel[b]

[a] *New Technologies Laboratory, INRETS (Institut National de Recherche sur les Transports et leur Sécurité), 94114 Arcueil, France*

[b] *LAAS (Laboratoire d'Automatique et d'Analyse des Systèmes), CNRS, 31077 Toulouse, France*

[c] *Presently at ABB Corporate Research Center, 5405 Baden-Dättwil, Switzerland*

**Abstract**

To study the failure mechanisms induced on high power IGBT multichip modules by thermal cycling stress in traction environment, a good knowledge of the temperature distribution and variations on the chips and in the interfaces between the different layers of the packaging is necessary. This paper presents a methodology for contact temperature measurements on chips surface in power cycling conditions and a fast 3D thermal simulation tool for multilayered hybrid or monolithic circuits. The results of static and dynamic thermal simulation of a 1200A-3300V IGBT module are given and compared with the contact temperature measurements results.

The investigation has been done within the RAPSDRA (Reliability of Advanced High Power Semiconductor Device for Traction Applications) European project. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. High power IGBT modules for traction applications and reliability investigations

In traction converters, semiconductors are submitted to harder working conditions than in the most power industrial applications. In particular, traction semiconductors are submitted to thermal cycling stress due to the great number of station halts, to the line profile, to the thermal external environment and to traffic. In case of guided urban transportation systems, thermal cycles in the semiconductors follow the traction current variations and their period is about one minute. The life expectancy required for the railway vehicles equipment is typically 30 years. This corresponds to approximately 5 million of inter-station thermal cycles and a working time of 100 000 hours.

The recently manufactured IGBT modules for traction applications reach very high powers (1200A-1600/3300V). Their technology uses wire bonding with a lot of high voltage chips soldered on ceramic substrate, which is new for traction semiconductors. Because of their technology, the IGBT modules are particularly exposed to thermal fatigue failures due to power cycling. Thus, reliability investigations become an important issue for the IGBT modules manufacturers as well as for the railway equipment industry.

Power cycling accelerated tests have been made at INRETS-LTN laboratory since 1993 to study the power thermal fatigue mechanisms and their influence on the life expectancy of high power IGBT modules [4]. We identified two principal failure mechanisms : one concerning a degradation of the bonding areas between the wires and the silicon chips, and the other concerning a degradation of the solder layers between the ceramic substrates and the copper base-plate.

Thermal characterization of the modules in power cycling working conditions is of great

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00134-6

importance for the reliability investigations. The classical junction temperature measurement by electrical parameters gives a single value representing nearly the average temperature of the silicone chips. Thus, a local temperature measurement in the real working conditions of the modules is necessary to evaluate the maximal thermal stress and the temperature gradient on chips surface. This way, we can also get the temperature values and variations near the bonding areas, which is a step forward in the understanding of the degradation process of the wire bondings. To get the full temperature distribution on the weak solder interfaces (between the ceramic substrate and the base-plate, for example), the only way is the simulation.

## 2. Description of the contact temperature measurements method chosen

Several methods can be used to measure the temperatures on chip surface : by Infra Red cameras after removing the silicone gel which covers the chips, by fine thermocouples or by optical fibers with crystal senses on their ends and which operate in contact with the chip surface. We chose this last method for the temperature measurements on chip surface in static and power cycling conditions because of its dynamic behavior and also because we can make the measurements without removing the silicone gel from the open modules. This way, the thermal environment of the silicon chips is unchanged and the temperature distribution obtained is nearly the same as in the real cycling conditions.

On the optical fibers we used, a temperature sensor consisting of a small amount of temperature sensitive phosphor is deposited on the end of an optical fiber probe. The sensor is excited with blue violet light pulses every 250ms and it exhibits a deep fluorescence. The fluorescence decay time is measured and correlated with the phosphor temperature.

Since the thin layer of phosphor is exposed without a protective outer coating and is bonded to the surface of the optical fiber with a soft elastomer, it makes good thermal contact with any surface it is pressed against without thermally perturbing the surface. As the elastomer deforms, it brings the phosphor layer into direct contact with the surface, thus insuring good thermal contact.

The sensing spot size of a probe is 0.25mm. The response time at 63% of these probes is about 25ms with a measurement every 250ms. The accuracy given by the manufacturer is better than ±1°C (measured on an aluminum block). The contact measurement accuracy on IGBT chips surface is about ±3°C.

## 3. Presentation of the thermal simulation tool LAASTHERM

As it is difficult to make contact temperature measurements on modules in industrial applications, it is useful to make thermal modeling of their packaging. In addition, the simulation can provide the temperature distribution at the interfaces between different layers, in particular the weak solder layers, which can't be done by measurements. In this context, a fast 3D thermal simulation tool for multilayered hybrid or monolithic circuits, developed by LAAS laboratory and named LAASTHERM, was used for temperature estimation in static and power cycling working conditions of a 1200A-3300V IGBT module.

LAASTHERM is a fast semi-analytical semi-numerical 3D thermal simulation tool for multilayered structures [2]. It is based on the use of the Two-Port Network Theory, together with simple mathematical transformations as Fourier transformation, in the case of 3D heat conduction problems. In other words, the Fourier integration functions are used to transform the heat equation into a simple differential equation. After that, the Two-Port Network Theory is applied to the transformed temperatures and fluxes. Numerical algorithms are only used to evaluate the solution and adapt it to the problem's geometry, which needs small computational effort. For example, the static calculations presented below took about 1 minute on a Pentium II (300MHz) machine while the dynamic simulations took 10 minutes.

## 4. Static thermal modeling of a 1200A-3300V IGBT module

A static thermal modeling was achieved on a 1200A-3300V IGBT module and compared with the temperature measurements results obtained on such module in the same working conditions.

### 4.1. The model description

The model is represented by a set of rectangular blocs (see fig. 1). The physical properties of the different materials and the contact resistance at the interfaces can be fixed by the user. The electrical power is applied on the dissipation surface of each silicon chip and the temperature of the lowest surface is fixed to the cooling system temperature.

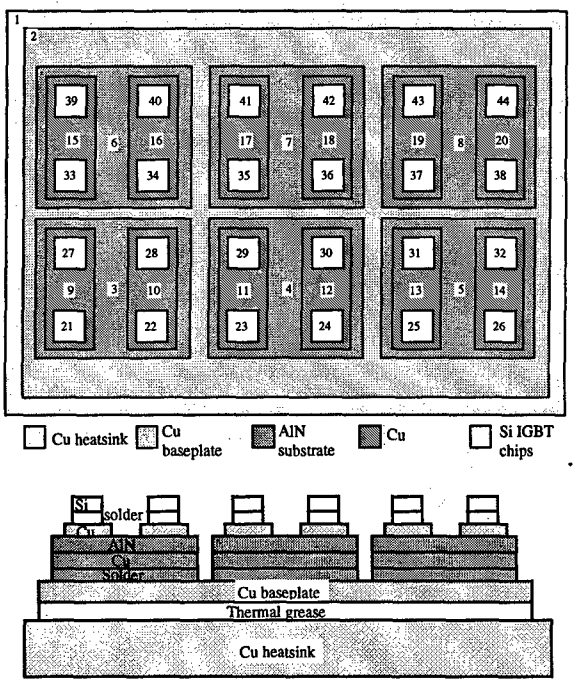



Fig. 1. Geometry of the IGBT module model

### 4.2. Working conditions applied to the model

We applied to the model the electrical and thermal conditions that were used in the experimental temperature measurements test. A total power of 5.6kW was dissipated in the module with a reference temperature on the bottom of the copper heat-sink of 38.5°C. This temperature corresponds to the average cooling water temperature in the experiment.

The contact thermal resistance between the copper heat-sink and the "water" at the reference temperature was fixed in the model to 0.2Kcm$^2$/W. A contact thermal resistance was also defined between the ceramic layer and the copper of the DCB substrate. It was fixed in the model to 0.04Kcm$^2$/W on each side of the ceramic layer.

### 4.3. The simulation results

The temperature distribution obtained on the chips surface by the simulation tool are presented on fig. 2 and 3 and are directly compared to the contact temperature measurements results. The junction temperature measured by the indirect electrical method (Vce measurement) is equal to 123°C.

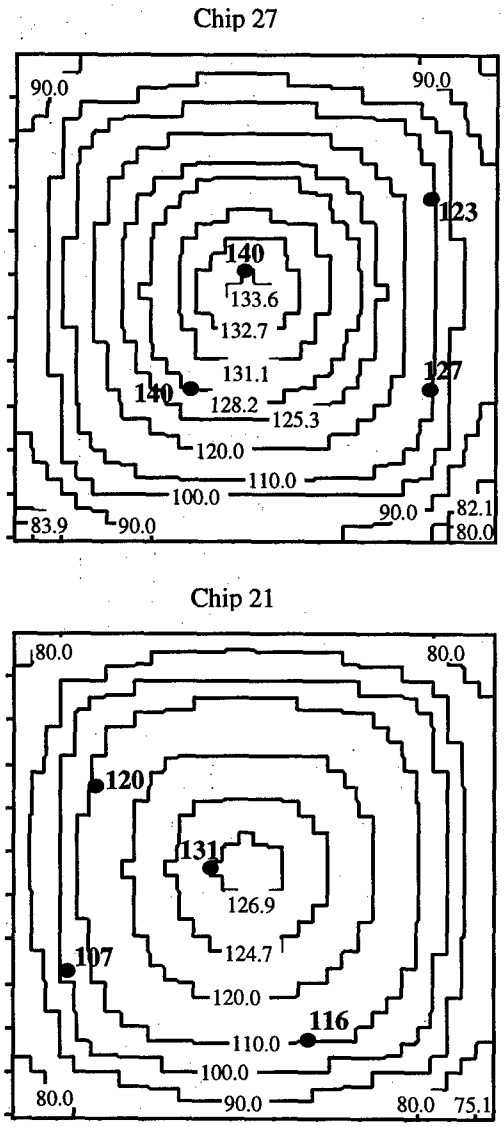



Fig. 2. Temperature distribution on chips 27 and 21 in static working conditions (233W/chip, reference temperature at 38.5°C). Simulation isotherms and measurement points (measured temperatures are in bold characters). Values given in °C.

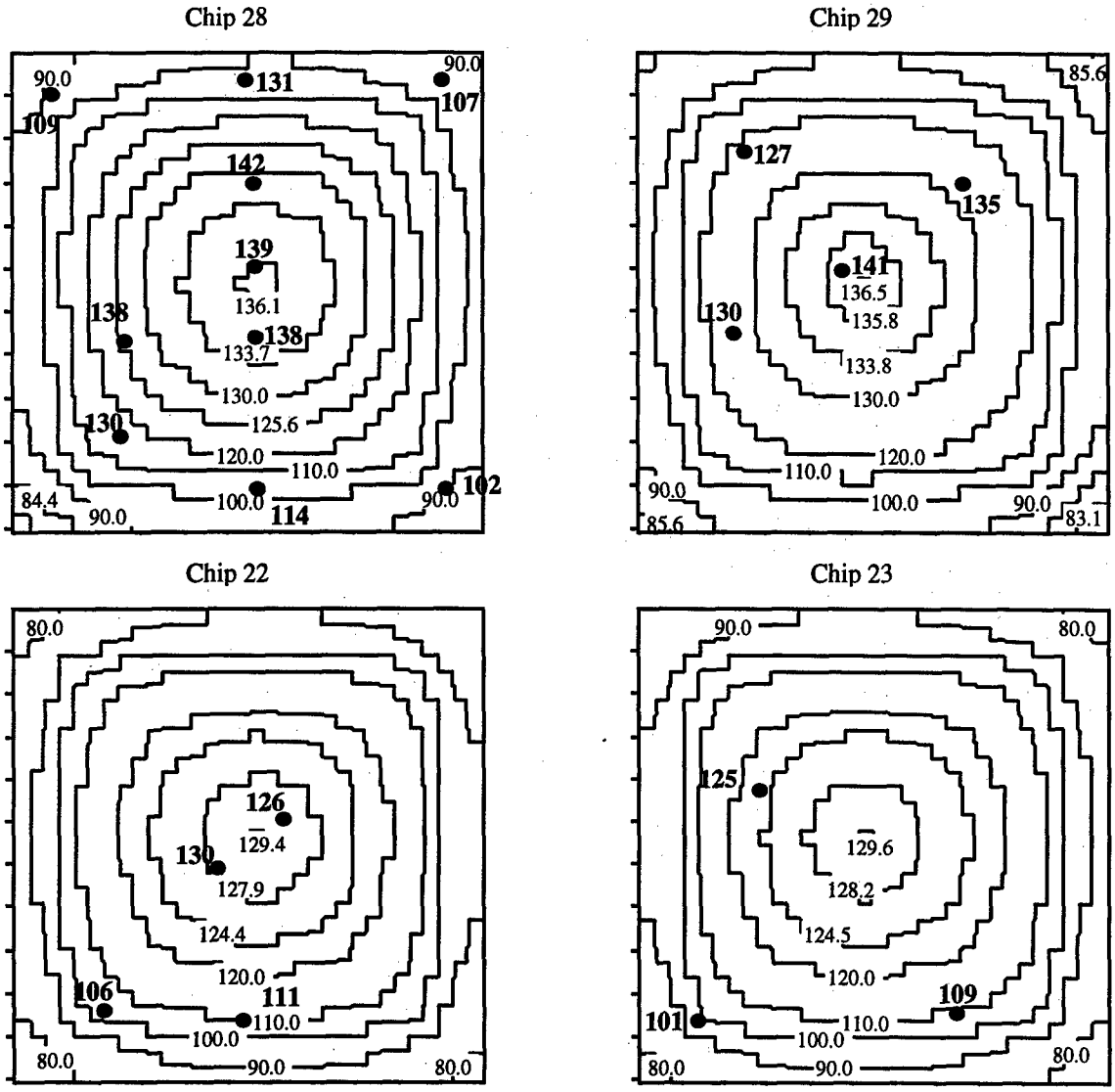



Fig. 3. Temperature distribution on chips 22, 23, 28 and 29 in static working conditions (233W/chip, reference temperature at 38.5°C). Simulation isotherms and measurement points (measured temperatures are in bold characters). Values given in °C.

We observe that the measured temperature values in the center of the chips are close to the simulation results while the difference between the simulation and the measurements is higher in the corners of the chips. This is due to the high temperature gradient at the corners of the chips and to the temperature sensors position error (around ±1mm). In addition, the calculation is made on a finite number of knots (32x32 on the Si chips, 64x64 on the ceramic substrate, 256x128 on the base-plate, grease and heat-sink layers).

On the other hand, it appears clearly that the difference between the measurements and the simulation results is more important on the chips 27 and 28. We can explain this error by the position of the chips 27 and 28 over an area where the thermal grease thickness is higher than the average thickness (see fig. 4). The simulation tool can't describe variable thickness of a layer yet.

In addition to that, errors can occur because of electrical or thermal differences between the different chips or because of local higher

temperatures near the bonding areas. All these effects are not taken into account in the simulation.

Figure 5 represents the calculated temperature distribution under the base-plate (interface copper base-plate/thermal grease) compared with contact temperature measurements results. On this layer, the difference between the measurements and simulation results is lower than on chips surface. That is probably due to the lower thermal gradient encountered.

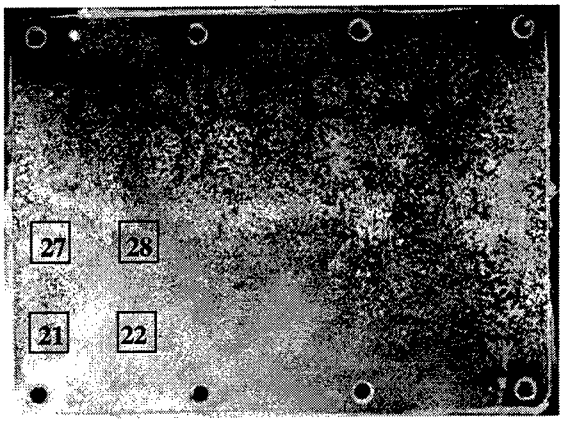



Fig. 4. Thermal grease distribution on the heat-sink after removing the module.

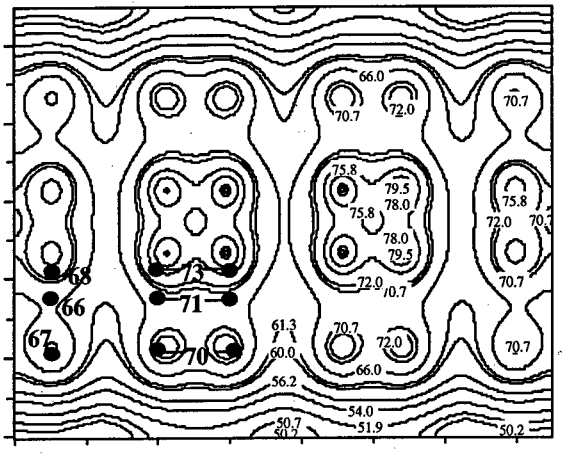



Fig. 5. Temperature distribution (in °C) at the interface base-plate / thermal grease in static conditions (233W/chip, reference temperature at 38.5°C). Simulation isotherms and measurement points (in bold characters).

## 5. Dynamic thermal modeling of a 1200A-3300V IGBT module in power cycling conditions

The dynamic simulation can be made only with a mono-bloc structure. Thus, the IGBT module was described by a set of layers having all the copper base-plate dimensions. The electrical power was applied on the dissipation areas on top of the silicon layer and the temperature of the lowest surface was fixed to the cooling system temperature.

To evaluate the errors induced by this different geometry, we applied on it the same conditions than in the static simulation and run a calculation with a heating time of 30s (to reach the steady state). We obtained a maximal error of -5°C on the dynamic model compared to the static one.

### *5.1. Working conditions applied to the model*

During the experiment, the module was submitted to power cycles of 30s on with permanent current / 30s off without current. The water cooling was applied continuously. The power applied during the heating phase was of 193W/chip. We used this value for the simulation.

The junction temperature measured by the indirect electrical method (Vce measurement) at the end of the heating phase was equal to 74°C.

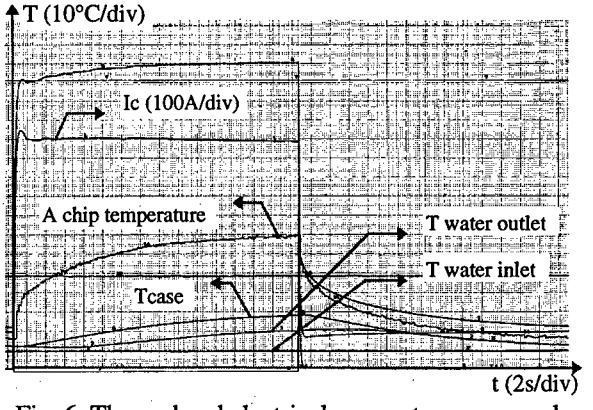



Fig. 6. Thermal and electrical parameters measured on the IGBT modules during a power cycle

We applied a linearly variable reference temperature (between 10.5°C and 14.5°C during the heating phase) under the heat-sink of the model (reference temperature), which corresponds to the average cooling water temperature.

A calculation was made each 200ms and there were 64x64 simulation knots on each layer.

*5.2. The simulation results*

The first simulation results showed faster temperature variations than the measurements resulting in high errors. To solve this problem, we changed the thermal grease thickness, which is not a well known parameter. We noticed that decreasing this parameter (e) results in better concordance between the simulation and the experiment (see fig. 7), we fixed it to 20μm which remains a reasonable value.

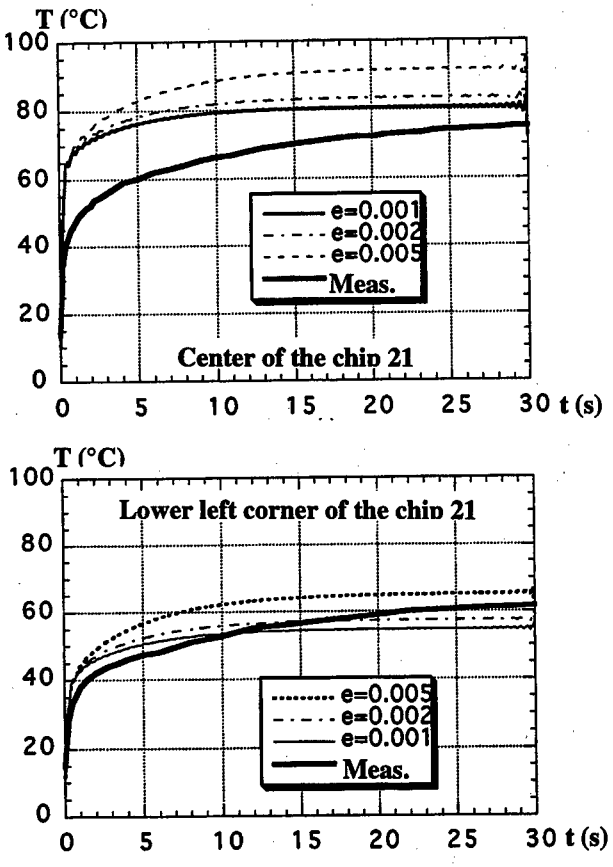



Fig. 7. Thermal grease thickness effect on calculated thermal profiles on chips surface

The dynamic simulation results are presented on figure 8. The simulation gives faster temperature variations than the experiments which results locally in high differences between the calculation and the experiment. These errors can be due to different parameters : a simplified description of the heat-sink and the water cooling system, errors on the thickness and the physical parameters of the different layers of the packaging which are very difficult to measure...

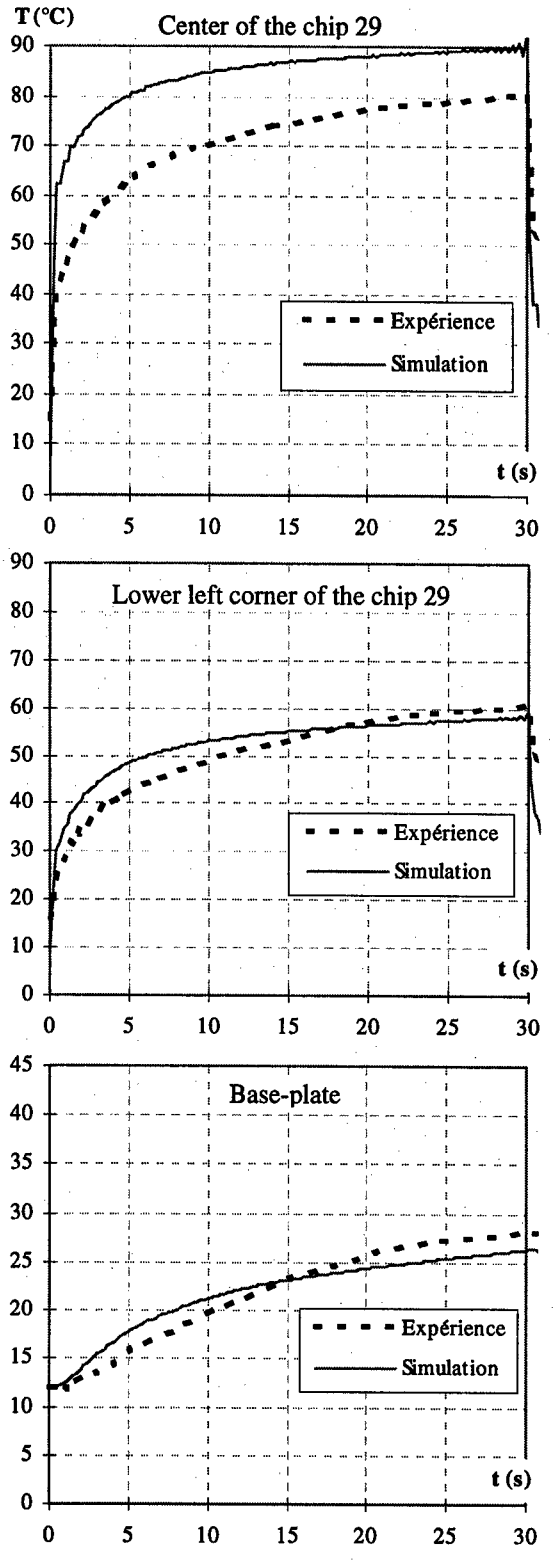



Fig. 8 : Calculated and measured temperature profiles during the heating phase of a power cycle.

## 6. Conclusion

Concerning the validity of the simulation, the analysis of the results are based on the comparison with the experimental values.

A static simulation was presented and compared with the temperature measurements results. Maximal errors around 3% on the center of the chips and 10% on the corners were generally noticed. Higher errors were observed on the chips situated over areas where the thermal grease is thicker. This problems can be solved by creating, within the simulation program, a procedure which takes into account thickness irregularities of the different layers.

A first dynamic simulation in power cycling conditions gave less performant results. Errors were noticed in the temperature variation slopes. This can be improved with a better knowledge of the modules packaging technology and the physical properties of their materials.

Concerning the physical interpretation of the data delivered by this simulation tool and the model used, we got two major results for the thermal fatigue investigation. The first one is that we obtained the thermal gradient on the soldering layer interface between the ceramic substrate and the base-plate which is a very important step forward in the thermal fatigue and the thermo-mechanical behavior investigation. The second result is about the big influence of the thermal interface between the module base-plate and the heat-sink. The grease layer is very difficult to model because of its relatively unknown thickness, its inhomogeneity over the whole interface area and its instability during the modules' ageing duration.

However, fast static and mainly dynamic simulation tools are necessary to get the thermal behavior of the high power modules accurately. From the reliability point of view, in particular concerning the thermal fatigue investigation, the final goal is to get the most accurate physical stress in working conditions.

## References

[1] Blackburn D.L. A review of thermal characterization of power transistors. 4th SEMITHERM symp. 1986.

[2] Dorkel J.M, Tounsi P and Leturcq P. 3D Thermal modeling based on the two-port network theory for hybrid or monolithic integrated power circuits. IEEE Trans. Components, Packaging and Manufacturing Technology, part A, vol. 19, n° 4, 1996, pp 501-507.

[3] Hamidi A and Coquery G. Effects of current density and chip temperature distribution on lifetime of high power IGBT modules in traction working conditions. ESREF, 1997, pp 1755-1758.

[4] Hamidi A, Coquery G and Lallemand R. Reliability of high power IGBT modules - Testing on thermal fatigue effects due to traction cycles. EPE, 1997, pp 3118-3123.

[5] Jacob P, Held M Scacco P and Wu W. Reliability testing and analysis of IGBT power semiconductor modules. IEE proc. IGBT propulsion drives, London, 1995, pp 4/1-4/5.

PERGAMON

Microelectronics Reliability 38 (1998) 1361–1366

MICROELECTRONICS RELIABILITY

# On-Chip Reliability Investigations on Power Modules Actually Working in Inverter Systems

T. Franke, M. Honsberg-Riedl, P. Simon*, J. Otto, S. Ramminger, G. Soelkner, E. Wolfgang

*Siemens AG, Corporate Technology, ZT KM 6, Munich*
** Siemens AG, Automation & Drives, MC E 42, Erlangen*

**Abstract**

The paper contains the results of temperature measurements for reliability investigations on IGBT modules actually working in inverters for industrial and traction applications. For this, the chip temperature and its transient behaviour have been monitored under different driving conditions of the inverters. Specially prepared IGBT-modules have been used to investigate the industrial and the traction inverters in their electrical and thermal behaviour. The measurement techniques were suitably adapted to the different configurations of the two inverters. © 1998 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

The market for power semiconductors and their applications is going through a phase of continuous growth thanks to the wider use of established products such as speed-controlled drives in manufacturing technology and the opening up of new applications such as transportation technology. Both of these require a continuous increase in system and component reliability.

The following three-step-strategy for assuring reliability is widely used to advantage for complex power electronic systems:

1) built-in reliability [1] by the chip and module manufacturers
2) reliability tests on the basis of standardized test conditions, where these are available
3) checking the reliability in the system.

The third point requires an exact understanding of the stress conditions occurring during operation. The temperature to which the components in an application are subject and its changes over time are regarded as the most significant parameters affecting the component's service life.

In the following investigations, these electrically-caused thermal characteristics of IGBT modules in inverters for variable-speed drives are measured and made accessible to a reliability evaluation. In this approach, modules are equipped with sensors for characteristic parameters, e.g. temperature, and are operated in the inverter under specified load conditions. The information obtained in this way allows a comparison to be made of the operational and test conditions and retro-actively permits components and systems to be designed for optimum reliability.

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00132-2

## 2. System based on-chip temperature monitoring

### 2.1. Consequences of reliability requirements

Industrial and transportation (traction) applications differ both with regard to their respective requirements on the system components and in the ease of obtaining statistical data from them.

Industrial inverters can be characterized as having a product life of about 10 years, high production quantities and numerous diverse applications.

Traction inverters are assumed to have a product life of 30 years, and this is reflected in the required failure rate of less than 100 FIT per module. Production quantities of these systems are small and their application range is well known.

For both, the enhancement of reliability comes with an enormous potential saving costs for service and failure handling for the costumer as well as for the manufacturer. The more accurate understanding of electro-thermo-mechanical processes inside the system components forces this enhancement.

### 2.2. Need for on-chip temperature monitoring

The relationship between a change in power dissipation on the chip and the expected temperature-time curve is shown schematically in Figure 1. The temperature cycles give rise to thermo-mechanical stress inside the module, which corresponds to the difference of the thermal expansion coefficients for the various materials used.

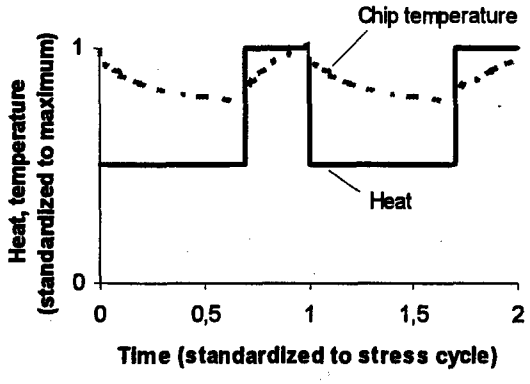



Figure 1. Temperature changes in modules caused by power cycles *(schematic)*

During the engineering process the transient thermal resistance values ($Z_{th}$) determined for the functional isolated components are commonly used to evaluate the chip temperatures of power semiconductors working in electronic systems.

But as a rule, the temperature fluctuations cannot be extrapolated exactly by the use of the $Z_{th}$-values [2] for real operating conditions of inverters. That is, because $Z_{th}$-values do reflect

- neither the change of actual amount of heat dissipation with altering driving conditions
- nor the hot spot but merely the average of chip temperature [2]
- nor the mutual thermal influence of different chips inside a module (e.g. the IGBT and diode chips or the six separate switches inside a fullbridge packaging)
- nor the dependency of the modules effective thermal resistance from the actual cooling system (e.g. air or water cooling).

Figure 2 illustrates the last two uncertainities.

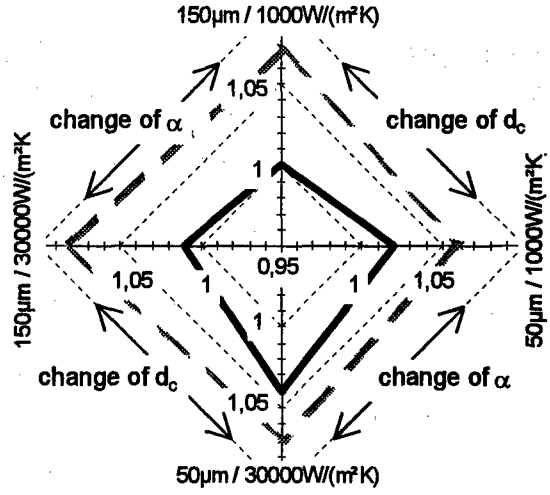



Figure 2. Normalized variablity of the thermal resistance ($R_{th,jc}$) of the central IGBT of a threepack with cooling: $d_c$ - thickness of thermo-lubricant between module and heat sink; $\alpha$ - heat transfer coefficient between heat sink and ambient and due to the mutual influence between different chips: straight line - only the central IGBT of a threepack produces heat; dashed line - all three IGBT of a threepack produce heat; *(finite element modeling)*

Therefore, the determination of the thermo-mechanical stress conditions caused by the load characteristics that affect the power semiconductors thus requires access to the chip temperature by measurement instruments.

## 3. Experimental set-up

### *3.1. Temperature sensing equipment*

Recent publications offer various methods for extracting chip temperature, such as monitoring of temperature dependent device characteristics itself, using infra-red cameras, contacting the chip surface with thermocouples or applying optical-fiber temperature sensors [4]. Only the optical methods should be used, however, for "hot" measurements in inverter systems to avoide hazards caused by electro-magnetic interaction between the device under test and the temperature probes.

#### *3.1.1. Infra-red camera*

Infra-red imaging is common practice to evaluate heat fluctuations and temperature distributions. The major advantage comes with the contactless temperature imaging which is responsible for an inherent capability for high frequency applications in the range of 1MHz. However the response of commercial systems is limited to 25Hz by the standards of video frequency.

During this work for infra-red imaging unencapsulated IGBT modules were used whose chip surface had been covered by a thin isolating resist layer (urethan) as well as a top layer of magnesia to uniform the emissivity [3].

#### *3.1.2 Optical-fiber sensor*

Currently different types of optical-fiber temperature measuring systems are available, such as phosphorous, GaAs and Fabry-Perot elements. Most of them come with the disadvantages of contact sensors. That means, that the response and the accuracy depend on the quality of the thermal contact between the surface and the sensor. Both are influenced by the thermal resistance of the non-ideal thermal contact and the heat capacitance of the sensor and any kind of thermo-lubricant between. Consequenly the contact sensor introduces an additional time constant in the measurement and represents furthermore an additional heat sink for short power pulses. By that the response of thermal contact measurement is restricted to less than 0.3kHz. The effect on the signal to measure and the recorded temperature are shown in Figure 3.

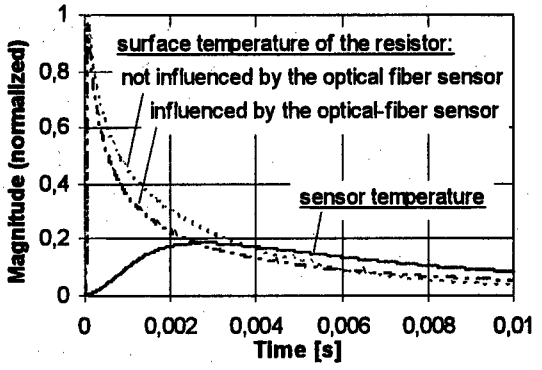



Figure 3. Influence of the optical-fiber sensor application on the signal to measure and the recorded data; the sensor temperature is the measured response to a deltapulse of heat dissipated in a resistor, the surface temperatures are results of calculations carried out with a 1D-transmission line model fitted to the actual experimental setup

The probe used consists of a Fabry-Perot element mounted at the end of an optical fiber [4], [5]. The Fabry-Perot element works like an interferrometer with temperature dependent index of refraction: The phase shift of a reflected light pulse is used as the measure for temperature.

For good response and accuracy the end facet of the sensor has been covered with a very small amount of thermo-lubricant and pressed against the surface. A glass tubing system was installed for guiding and fixing the optical-fiber sensor.

### *3.2 Power electronic*

The time versus temperature curves of several IGBT and diode chips were determined on two inverters from different application sectors and different power classes:

- industrial inverter for approximately 25kW switched power based on 1200V-class threepacks and
- traction inverter for approximately 500kW switched power based on 3300V-class single switches.

*3.2.1 Industrial inverter for approximately 25kW*

The electrical layout for an approximately 500V link-circuit voltage and 50A motor current permitted the unencapsulated preparation of the power semiconductors and their observation during operation with an infra-red camera. Additionally, for fast transient temperature measurement the optical-fiber system was used [4], [5]. The preparation of the module had been carried out as described in section 3.1. Unfortunately, during this measurements the response of the optical-fiber sensor was limited to the range of 50Hz due to the additional urethan layer which had to be used to prevent electrical hazards from the unencapsulated module.

*3.2.2 Traction inverter for approximately 500kW*

The electrical layout for an approximately 2kV link-circuit voltage and 500A motor current requires the closed preparation of the power semiconductors in order to maintain the operational reliability of the inverter. Therefore only the optical-fiber system has been used to record the temperature of selected chips. The preparation of the module had been carried out as described in section 3.1.2. The full band width of the optical-fiber sensor (0.3kHz) was available because of the direct application of the sensors to the chips without any isolating coating.

## 4. Experimental Results

*4.1. Measurement uncertainty*

Both measurement techniques were compared on the basis of similar laboratory measurements in the 25kW industrial inverter by using standard thermal-characterization procedures with $Z_{th}$-measurement [2]. Their results were also compared with those of the simulation, that means time independent calculations by finite element modeling. The static measurement uncertainty was found to be in the range of 0.5K or 3% of the full magnitude. For transient uncertainty see section 3.1. and figure 3.

*4.2. Industrial Inverter*

Due to the possibility of unencapsulated preparation and the easier handling of low power from the industrial inverter more detailed data could be obtained.

Different stress cycles were specified in the experiments to reflect the requirements of industrial practice. The results of three measurements are given as an illustration.

Figure 4 shows the response of the IGBT temperature to a short increase (0.2-second) of the rated motor current.

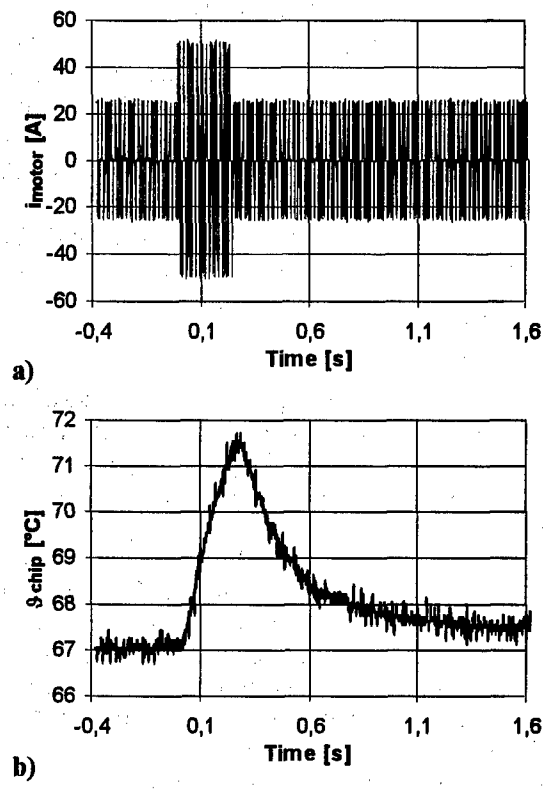



Figure 4. Temperature curve of an IGBT chip with a 0.2-second increase in the rated motor moment from 70% to 130% at a cycle time of 10 seconds: a) current of one motor phase; b) temperature of the switching IGBT

Figure 5 shows the temperature response for a load cycle with a 4-minute increase in the rated motor moment. Two processes can be seen with different time constants. They can be assigned to the module (approx. 1s) and the heat sink (approx. 20min), respectively.

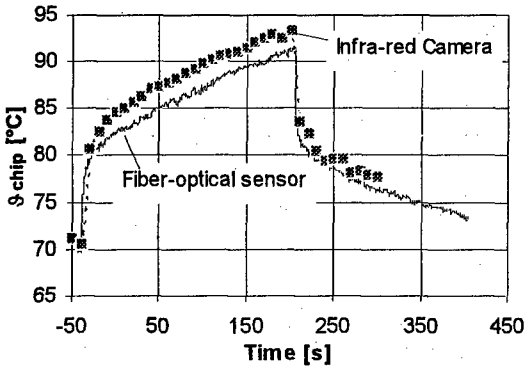



Figure 5. Temperature curve of an IGBT chip with a 4-minute increase in the rated motor moment from 70% to 130% at a cycle time of 10 minutes

Figure 6 shows the characteristic states of a reversing process. It was shown that the highest stress due to temperature change occurs at low speeds (in the region of the reversal of direction). The maximum chip temperature in this case directly follows the current of the associated motor phase.

*4.2. Traction Inverter*

The possibility enlarging power capability of IGBT-modules by paralleling several chips does not come without the problem of inhomogeneous thermal environment across the modules area. Problems in system engineering are the homogeneous thermal contact of the module to the heat sink and the shift of heat sink temperature itself along the way of air or water flow. For detailed understanding the tested IGBT module has been equipped with six optical-fiber sensors placed on the top of three IGBT and three diode chips. For calculation of the effective heat flow six Pt100-type temperature sensors have been placed into the module base plate underneath the observed chips.

Figure 7 shows the distribution of chip and case temperatures for the IGBT module related to various electrical working conditions. During different measurements the dissipated power as well as the maximum chip temperature have been adjusted to the same range avoiding nonlinear semiconductor effects in heat dissipation. Thermal conditions have not been altered. The analysis of the results shows the redistribution of temperature with altering working conditions due to the mutual influence between IGBT and diode chips.

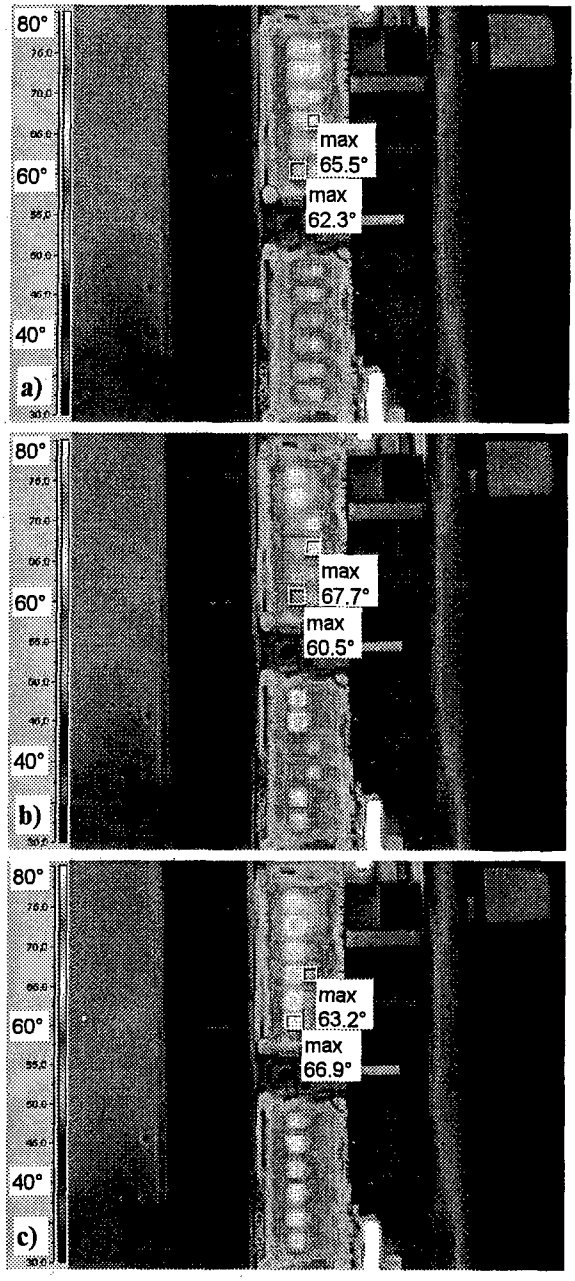



Figure 6. False-color representation of the surface temperature of an uncovered IGBT inverter undergoing a reversing process: a) slowing, the stress shifts from the IGBTs to the diodes; b) change of direction - the chip temperature directly follows the current of the motor phase; c) acceleration, the stress shifts to the IGBTs

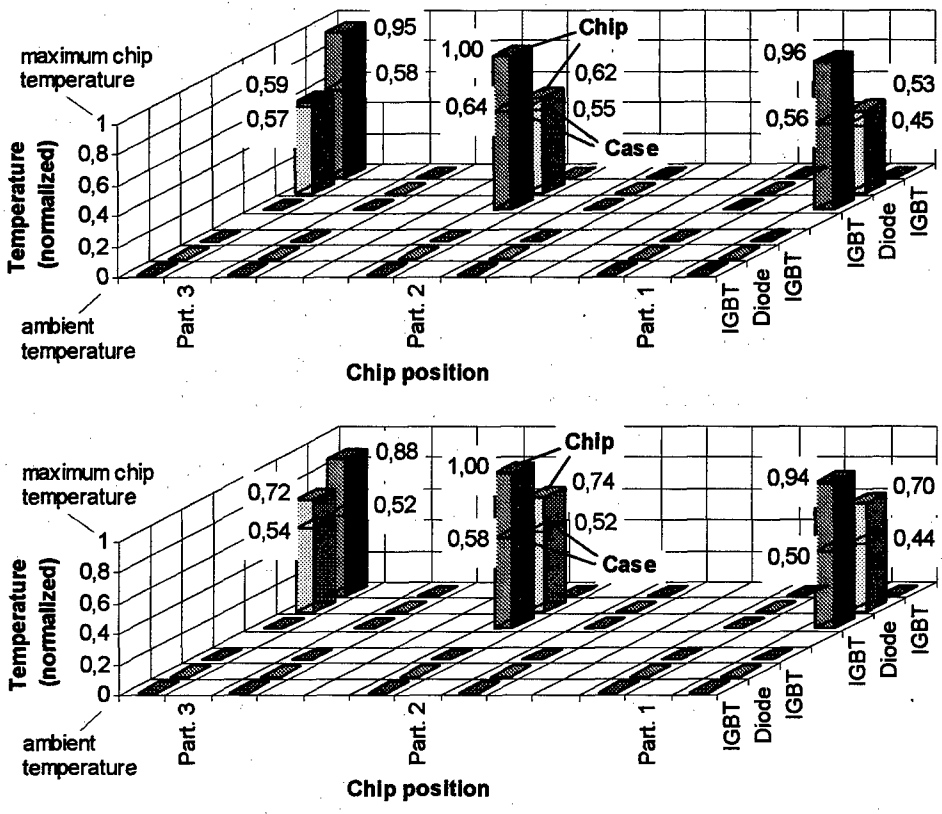



Figure 7. Temperature distribution on chips and case across a 3.3kV-IGBT-module during operation: a) only the IGBT chips of the whole module dissipate power *($Z_{th}$-measurement)*; b) IGBT and diode chips dissipate power at the same current level *(inverter configuration with inductive load)*; *(temperatures are normalized to the maximum deviation from ambient temperature)*

## Conclusion

IGBT modules have a central function in power electronic equipment for speed-controlled drives. The understanding of the interaction between the semiconductors and the inverter system is important to reach high reliability. This work presents how the main important parameter for characterizing stress conditions - the chip temperature - can be measured during real inverter operation at full power. The results presented illustrate the temperature cycles coming of transient as well as static inverter operation and show the variability of temperature distribution by the mutual influence between several chips. The development of a model describing the mutual influence based on measured parameters is currently in progress.

## Acknowledgement

The authors thank K.-H. Sommer and T. Nuebel (eupec) as well as G. Zaiser and T. Weigel for experimental assistance.

## References

[1] Crook, D. L.; "Evolution of VLSI Reliability Engineering", Proc. of IRPS 1990, pp. 2-11

[2] Lappe, R.; Fischer, F.: Leistungselektronik - Meßtechnik; Verlag Technik GmbH, Berlin 1993

[3] Poech, M.-H.; "Preparation for Infrared Imaging at 8-12µm", personal reference 1995

[4] Hamidi, A.; Coquery, G.; Lallemand, R.; "Reliability of High Power IGBT-Modules Testing and Thermal Fatigue Effects due to Traction Cycles", Proc. EPE '97, Sept. 1997

[5] MetriCor; "Fiber Optic EMC Test System METRICOR 1420", Operators Manual, MetriCor, Wakefield Nov. 1997

PERGAMON

Microelectronics Reliability 38 (1998) I–II

MICROELECTRONICS RELIABILITY

# Author Index

Alba, S., 919
Al-Kofahi, I. S., 1085
Arnaud, L., 1029, 1041
Arzt, E., 1015
Auriel, G., 937
Aymerich, X., 1127

Bae, S., 969
Baiget, F., 937
Balboni, R., 1021
Baldi, L., 919
Balestra, F., 1103
Balk, L. J., 957, 963
Bangert, J., 951
Barrett, J., 1277
Beech, C., 1085
Behnke, U., 981
Benbrik, J., 901, 1265
Bender, H., 877
Benteo, B., 901, 907, 1265
Berg, H., 1319
Bock, K., 997
Boivin, P., 937
Bonfiglio, A., 1215
Bonnaud, O., 1149
Borgarino, M., 1199, 1233
Borowski, J., 1171
Bouwstra, S., 1251
Bouzid, M. J., 1115
Braun, P., 1245
Brederlow, R., 1063
Brini, J., 1057
Buiu, O., 1085
Burmer, C., 987

Carlsson, S.-B., 943
Casanovas, V., 1171
Casu, M. B., 1215
Cattani, L., 1199, 1233
Cavernes, P., 1171
de Ceuninck, W., 1009, 1035, 1187
Charpenel, P., 1171
Chen, M. S., 1287
Chin, R., 963
Chopin, J. M., 1171
Christensen, C., 1251
Christou, A., 1239, 1307
Ciappa, M., 889
Claeys, W., 1293, 1341
Coquery, G., 1353
Cornet, A., 1293, 1341
Courtade, F., 907
Cova, P., 1199, 1347
Cramer, R. M., 963
Croes, K., 1187
Crosato, C., 1227

Danto, Y., 901
Das, A. G. M., 1177
Deane, P., 1183
Defars, J.-P., 1221
Deleonibus, S., 1103
Deppert, K., 943
Desplats, R., 901, 1265
Devoldère, P., 1221
D'Haeger, V., 1009
D'Haen, J., 1035
Dilhaire, S., 1293, 1341
Dimoulas, A., 1307
Dorkel, J. M., 1353
Duane, R., 1091
Dyrbye, K., 1251

Esser, R. H., 1307
Estève, D., 1261, 1265
Etoh, Y., 1313

Fantini, F., 1021, 1233, 1347
Feige, V., 957
Feng, T., 1239
Fiege, G. B. M., 957
Fily, A., 1245
Findeisen, C., 1335
Foley, N., 1121
Foley, S., 1021
Franke, T., 1361
Frémont, H., 1261
Friis Eriksen, G., 1251
Fujioka, H., 975
Fukuda, M., 1211
Fukumoto, K., 913
Furcas, P., 889

Galateanu, L., 1331
Garat, F., 1227
Gauffier, J-L., 907
Ghibaudo, G., 1057, 1075, 1103
Glachant, A., 937
Gobil, Y., 1029
Golze, U., 1193
Gonella, R., 1029
Görlich, S., 957, 987
Gounelle, C., 1029
Gravesen, P., 1251
Groeseneken, G., 877, 997
Guégan, G., 1103

Haddab, Y., 1069
Hamidi, A., 1353
Hansen, C. K., 1155
Härtler, G., 1193
Hashikawa, N., 913
Heiderhoff, R., 963
Herr, E., 1335
Hoffmann, K., 1139
Honsberg-Riedl, M., 1361
van Hove, H., 1097
Hurley, P. K., 931
Hussain, S. Z., 851

Iga, R., 1211
Ikeno, M., 913
Impronta, M., 1021
Ingvarson, F., 1109
Inoue, S., 883
Itaya, Y., 1211
Ito, S., 993

Jansson, P., 1297
Janting, J., 1251
Januszewski, S., 1325
Joannes, L., 1341
Johnson, S., 1177
Junno, T., 943
Jurk, R., 1063

Kayama, T., 1313
de Keukeleire, C., 1097
Kies, R., 1057
Kim, D. W., 1081
Kociszewska-Szczerbik, M., 1325
Kowarik, O., 1139
Kraft, O., 1015
Krog, J. P., 1251
Kubalek, E., 951, 969
Kuper, F. G., 1287
Kuroi, T., 913
Kwang, JL., 907

Labat, N., 901
Lallemand, R., 1353
Landesman, J. P., 1245
Larsen, A., 1271

0026-2714/98/$ – see front matter. © 1998 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(98)00168-1

Lauwers, A., 925
Lazzarini, L., 1199
Lecouvé, M., 1261
Lee, B. J., 1081
Lekens, G., 1035
Lele, A., 1183
Liang, Z. N., 1287
Linnenbank, C. G., 1063
Liverød, B., 1271
Lloyd, J. R., 851
Lormand, G., 1041
Lundgren, P., 1109

Maes, H. E., 877, 997
Maex, K., 1009, 1035
Magistrali, F., 1215
Magnusson, M. H., 943
Malberti, P., 889
Manca, J. V., 1009, 1187
Manic, D., 1069
Marchand, B., 1103
Martin, A., 1091
Martin, E., 1245
Martines, G., 1227
Mashiko, Y., 913
Mathewson, A., 931, 1021, 1091
Maywald, M., 957, 969
Mazzer, M., 1199
Meneghesso, G., 1227
Mertin, W., 969, 981
Meunier, D., 1265
Mino Diaz, E., 1165
Miranda, E., 1127
Miura, K., 975
Mohammed-Brahim, T., 1149
Molenberghs, G., 1187
Mondon, F., 1075
Monma, H., 993
Montelius, L., 943
Morand, Y., 1029
Morin, M., 1221
Mourgues, K., 1149
Mourrain, Ch., 1115
Mouthaan, A. J., 1047
Mühle, U., 895
de Munari, I., 1021

Nafría, M., 1127
Nakamae, K., 975
Nassim, K., 1341
Nikawa, K., 883
Nishiya, T., 1211
Nysæther, J. B., 1271

Ohlckers, P., 1271
Ohtsu, M., 883
van Olmen, J., 1009
O'Sullivan, P., 1091
Otto, J., 1361
Oualid, J., 937
Overhauser, D., 851

Paccagnella, A., 919, 1227
de Palo, R., 1215
Pananakakis, G., 1057, 1075
Pantisano, L., 919
Papanicolau, N., 1307
Park, J. T., 1081
Paschke, K., 1193
Pauthner, S., 987
de Pauw, P., 877
Pellet, C., 1261, 1265
Perdu, P., 901
Perez, G., 907, 1265
Petrescu, V., 1047
Pettarin, L., 919
Phang, J. C. H., 957
Pichon, L., 1149
van der Pol, J. A., 1051
Popa, E., 1331
Popovic, R. S., 1069
Prendergast, J., 1121

Radjaa, S., 937
Ragnarsson, L-Å., 1109
Rahal, A., 1149
Ramminger, S., 1301, 1361
Raoult, F., 1149
Rasras, M., 877
Rauzan, C., 1293
Reczek, W., 1139
de Reus, R., 1251
Riess, P., 1057
Rochel, S., 851
Rodríguez, R., 1127
Romedahl Brown, T., 1251

Sagnes, B., 937
Saiki, T., 883
Salm, C., 1047
Salmini, G., 1215
Salviati, G., 1199
de Salvo, B., 1075
Samuelson, L., 943
Sasaki, K., 1313
Scarpa, A., 919
Schade, W. R., 963
Schaub, E., 1293, 1341
Schenkel, M., 1335
de Schepper, L., 1009, 1035, 1187
Schlegel, R., 1335
Schleifer, H., 1139
Schlensog, A., 969
Schlünder, C., 1063
Schmitt-Landsiedel, D., 1063
Schoenmaker, W., 1047
Schray, S., 895
Scorzoni, A., 1021
Setti, G., 1143
Sheehan, E., 931
Simon, P., 1361
Soelkner, G., 1361
Søndergård Jensen, O., 1251

Speciale, N., 1143
Spolenak, R., 1015
Steele, G., 851
Stoica, M. G., 1331
Strifas, N., 1239, 1307
Suehle, J., 1121
Sugo, M., 1211
Suñé, J., 1127
Świątek, H., 1325

Takeshita, T., 1211
Tala-Ighil, B., 1149
Tartavel, G., 1029, 1041
Taylor, S., 1085
Thewes, R., 1063
Thomas, R. W., 861
Thyregod, P., 1155
Torrès, J., 1029
Touboul, A., 901
Tourniol, Ch., 1115
Toutah, H., 1149
Tsay, C., 1183
Türkes, P., 1301

Uffmann, D., 1133

Val, A., 1261
Valentini, G., 919
Vales, P., 1353
Vandamme, E. P., 925
Vandamme, L. K. J., 925
Vandenbossche, E., 1097
Vanhaeverbeke, S., 877
Vanzi, M., 889, 1215
Verkleij, D., 869
Vila Aresté, J. E., 1165

Wachutka, G., 1301
Walter, G. H., 1063
Waltz, P., 1041
Wand, B., 981
Wang, A. Z., 1183
Weber, W., 1063
Weichel, S., 1251
Wiesner, A., 895
Witters, J., 1097
Wittpahl, V., 981
Witvrouw, A., 1009, 1035
Wolbert, P. B. M., 1051
de Wolf, I., 877, 925
de Wolf, M., 1097
Wolfgang, E., 1319, 1361

Xu, H., 943

Yckache, K., 937
Yu, C. G., 1081
Yu, H. K., 1081

Zanoni, E., 1227
Zanotti-Fregonara, C., 1199
Zeller, H. R., 1335

# 1999 *INTERNATIONAL RELIABILITY PHYSICS SYMPOSIUM*

**March 22 – 25, 1999 • Town & Country Hotel • San Diego, California**

IEEE

## CALL FOR PAPERS

The International Reliability Physics Symposium deals with physical mechanisms that reduce the reliability or performance of integrated circuits and microelectronic devices in the users' environment. The conference is a forum for presenting original work that: (a) Identifies new microelectronic failure or degradation mechanisms, (b) Improves understanding of existing failure mechanisms, (c) Demonstrates new or innovative analytical techniques, or (d) Demonstrates ways to build in reliability. The conference aims to stimulate progress in the following broad areas:

- *Reliability Implications of Transistor Scaling*
- *Modeling and Simulation of Failure Mechanisms and Circuit Reliability*
- *Correlation of Reliability to the Physical and Electrical Properties of Microelectronics Materials*
- *Reliability-Driven Circuit Design and Wafer Processing*

## YOUR PAPERS ARE SOLICITED ON:

**Failure and Yield Enhancement Analysis**
–New or Innovative Analytical Techniques
–Analysis of New Failure Mechanisms

**Interconnect**
– Electromigration Phenomena in Al & Cu Conductors
– Microstructure and Processing Effects
– Mechanical Stress Related Reliability Issues
(film/line stresses, stress-voiding, hillocking, adhesion )
– Reliability of Cu / Low -k dielectric systems
– Reliability of Advanced Al metallizations
– Interlevel Dielectrics

**Device Dielectrics**
– Oxide Breakdown Mechanisms and Models
– Processing Effects on Gate Oxide Reliability
– Reliability of Ultra-Thin Gate Oxides (<50Å)
– Reliability of Alternative Gate Dielectrics
– Tunnel Oxides in Non-Volatile Memories

**Process Induced Damage**
– Latent Damage Characterization
– Damage Mechanisms in Advanced Technologies
– Reliability Degradation Associated with Damage

**Channel Hot Carriers**
– Oxide Degradation Mechanisms
– Modeling and Measurement Techniques
– Novel Issues in Low Power Applications

**ESD and Latch-Up**
– New or Innovative Structures
– Damage Interpretation
– Modeling and Measurement Techniques
– Analysis of ESD Failures

**Assembly and Packaging**
– Package Analysis Techniques
– Stress Modeling and Characterization
– BGA, C4 Package Related Reliability Issues
– Bonding Issues: TAB, Stud Bumping, DCA, etc..
– Multi-Chip Module, Printed Circuit Board, Connectors

**Non-Silicon Device Reliability**
– GaAs Devices
– Micro-Electro Mechanical Systems (MEMs)
– LEDs & Diode Lasers
– Flat Panel Displays

**Device and Process**
– Correlation Between Yield and Infant Mortality
– Reliability Effects of Particulate Control
– Soft Error Single Event Upset Mechanisms
– Examples of Building in Reliability
– Novel Early Failure Mechanisms in ULSI Circuits

**SUBMISSION DEADLINE:**
**September 11, 1998**

**Abstract Preparation:** Paper acceptance is based entirely on abstract submissions. Your work must be *original* and *unpublished.* Abstracts are a maximum of two pages long, and should clearly and concisely state the specific results of your work, why it is important, and how it relates to prior work. The abstract may include line drawings, key references and photographs as necessary, within the two page limit. Additionally, we require a 50-word summary of your work and a cover page indicating the category of submission from the above listing, as well as each authors affiliation, address, phone and FAX numbers and e-mail address.

**Abstract Submission:** Submissions may be made either by post , express mail, or electronically. All submissions will be acknowledged within three weeks of submission. If you do not receive acknowledgment of your submission, please contact the Technical Program Chair.

**Post or Express Mail:** We need 20 copies of both your 50-word summary and 2-page abstract.

**Mail to:** Anthony S. Oates, Technical Program Chair, 1999 IRPS
Bell Laboratories
Lucent Technologies
9333 John Young Parkway
Orlando, FL 32819, USA

Tel: 40/-345-7536
FAX: 407-345-7639
e-mail: a.s.oates@lucent.com

-over-

**Electronic Submission:** Abstracts will be accepted in electronic format. There are some restrictions on the file formats we accept. Please consult the following Web page for details of our electronic submission process, http://www.irps.org/paper_es.htm.

**Paper Selection:** Papers received by the September 11 deadline are reviewed by subcommittees representing the major sessions of the conference. The members of these groups are chosen from industry, government, and university organizations who have a broad spectrum of reliability experience and specific expertise in the areas of review. Papers are selected on their technical merits and relevance to the field of Reliability Physics. Some common causes of rejection are incomplete data, prior publication, or lack of novelty.

**Late Papers:** A limited number of excellent late papers reflecting important breakthrough developments can be considered on a space-available basis. Abstract and summary must be received ***no later than November 22, 1998*** to be considered. Late papers must still meet the publication deadline stated below.

**PROCEEDINGS MANUSCRIPT:** Final, camera-ready manuscripts must be received by ***January 25, 1999*** so that the proceedings can be available at the Symposium.

**SPONSORS:** The Electron Devices Society and the Reliability Society of The Institute of Electrical and Electronic Engineers, Inc. are the sponsors of the 1999 Symposium.




## For general conference information, contact:

USA
Alan G. Street
General Chair, 1998 IRPS
Integrated Reliability
10225 Barnes Canyon Road
Suite A-100
San Diego, CA 92121
USA
Tel. (619)-658-9770
FAX (619)-550-9445
e-mail: alan@irsi.com

for registration and mailings
IRPS Publishing Services
P.O. Box 308
Westmoreland, NY 13490
USA
Tel. (315) 339-3971
FAX (315) 336-9134
e-mail: 103227.2074@compuserve.com

Asia
Dr. Eiji Takeda
IRPS Publicity Committee
Hitachi Ltd.
PO Box 2
Kokubunji, Tokyo 185
Japan
Tel. (81) 423-23-1111 x3300
FAX (81) 423-27-7699
e-mail: ETakeda@CRL.Hitiachi.Co.Jp

Prof. Daniel S.H. Chan
IRPS Publicity Committee
Dept. of Electrical Engineering
National University of Singapore
10 Kent Ridge Crest
Singapore 119260
Tel (65) 772-6509
FAX (65) 779-1103
e-mail: elecshd@nus.edu.sg

Europe
Mauro Ciappa
IRPS Publicity Committee
Swiss Federal Institute of Technology
Integrated Systems Laboratory
ETH-Zentrum
CH-8092 Zurich, Switzerland
Tel. (41) 1-632-2436
FAX (41) 1-632-1194
e-mail: Ciappa@iis.ee.ethz.ch

Dr. Wolfgang Gerling
IRPS Publicity Committee
Siemens AG
P.O. Box 801709
D-81617 Munich
Germany
Tel. (49) 89 636 22825
FAX (49) 89 636 22828
e-mail:
WOLFGANG.GERLING@hl.siemens.de

# Call for Papers

## Special Issue of *Microelectronics Reliability* on

# Electromigration

The control of electromigration phenomena in interconnect systems is one of the cornerstones of IC technology development and design. Hundreds of literature articles are published every year which include reports of electromigration studies. The microelectronics industry is faced with the dual challenge of further understanding the physics of electromigration, while continuing to develop processes and design products which are immune to its effects. *Microelectronics Reliability* will be dedicating a special issue to electromigration, its history, physics, characterization, and performance optimization. Our goal is to offer a comprehensive overview of state-of-the-art developments in the field, with articles both informative to the newcomer, and novel to the expert.

Quality review and research papers are solicited in the following areas:

- Fundamental understanding of the physics of electromigration
- History of the study of electromigration phenomena
- Advances in modelling, simulation and test methodologies
- Applications of accelerated electromigration testing to IC product reliability
- Electromigration design rule specification in advanced IC designs
- Specific issues such as interconnect via electromigration lifetime
- Electromigration performance optimization of novel interconnect systems (Copper/low-k dielectrics...)
- Impact of other metal migration phenomena such as thermal and stress migration
- Future directions in electromigration research and engineering

## Submission Deadline: 31 October 1998

Please submit **four copies** of the manuscript to one of the following three guest editors:

## Special Issue Editors


C.D. Graas
ITTM
Siemens Microelectronics
1167 Route 52
Fishkill, NY 12524
USA
Tel: +1 914 897 1139
Fax: +1 914 897 5713
carole.graas@hl.siemens.de

A.J. Mouthaan
Faculty of Electrical Eng.
University of Twente
P.O. Box 217
7500 AE Enschede
THE NETHERLANDS
Tel: +31 53 489 2726
Fax: +31 53 489 2799
A.J.Mouthaan@el.utwente.nl

H. Onoda
VLSI R&D Center
550-1, Higashi-asakawa
Hachioji
JAPAN
Tel: +81 426 62 6107
Fax: +81 426 67 8367
honoda@ed.oki.co.jp


## Anticipated Publication Date: August 1999

# Call for Papers

## Special Issue of *Microelectronics Reliability* on

# Reliability of Compound Semiconductor Devices and ICs

The goal of this special issue is to review recent activities in the reliability physics and study of compound semiconductor devices and ICs, including optoelectronic ICs and other integrated optoelectronic devices, to update to current status, and to consider new future directions for such a subject. Papers on all aspects of compound material-based devices and ICs, such as physics, modelling, simulation, characterization and packaging are solicited. Innovative approaches and concepts for device, module, and circuit reliability prediction, testing, and improvement are particularly welcome.

### Topics

- Fabrication technologies
- Measurements of reliability performance
- Characterization and study of failure mechanisms and causes
- Reliability modelling and simulation
- Device design issue and innovative device concepts
- Packaging and module reliability

### Devices

- GaAs-based HBTs, HEMTs and MESFETs
- InP-based HBTs, HEMTs and MESFETs
- SiGe-based HBTs and HEMTs
- SiC-based MOSFETs and HBTs
- Lasers, LEDs, photodiodes and modulators
- Planar lightwave circuits and optical connectors
- Optoelectronic devices and ICs

## Submission Deadline: 30 November 1998

Please submit **four copies** of the manuscript to one of the following three guest editors:

## Special Issue Editors


Juin J. Liou
University of Central Florida
College of Engineering
Dept. of Electrical Eng.
Orlando, FL 32816-2450
USA

Fax : +1 407 823 5835
jli@ece.engr.ucf.edu

Fausto Fantini
Università di Modena
Dipartimento di Scienze
Via Vivaldi 70
41100 Modena
ITALY
Tel : +39 59 37 40 57
Fax : +39 59 36 41 32
fantini@dsi.unimo.it

Mitsuo Fukuda
NTT Optoelectronics Labs.
3-1, Morinosato Wakamiya
Atsugi-shi
Kanagawa, 243-01
JAPAN

mfukuda@aecl.ntt.co.jp


## Anticipated Publication Date: September 1999

NIŠ

MIEL '99


**STEERING COMMITTEE**

**Asia & Pacific**

S. Dimitrijev
*Griffith University, Australia*
H. Iwai
*Toshiba, Japan*
J. Kuo
*National Taiwan University, Taiwan*
C. Ling
*National University of Singapore*
Y. Luo
*Tsinghua University, China*
T. Ohmi
*Tohoku University, Japan*
S. C. Sun
*Taiwan Semiconductor Mfg.Co., Taiwan*
Y. Uchida,
*Fuji Electric Co., Japan*
H. Wong
*City University of Hong Kong*

**Europe**

B. Courtois
*TIMA-CMP, France*
H. Detter
*Technical University of Vienna, Austria*
V. Liberali
*University of Pavia, Italy*
M. Ostling
*Royal Institute of Technology, Sweden*
R. Popović
*EPFL, Switzerland*
P. Rossel
*CNRS-Motorola, France*
S. Selberherr
*Technical University of Vienna, Austria*
N. Stojadinović
*University of Niš, Yugoslavia*
V. Vashchenko
*PULSAR, Russia*

**America**

V. Arora
*Wilkes University, PA*
J. Clemens
*Lucent Technologies, NJ*
J. Colinge
*University of California at Davis, CA*
J. Liou
*University of Central Florida, FL*
A. Ortiz-Conde
*University Simon Bolivar, Venezuela*
T. Pearsall
*University of Washington, WA*
A. Salama
*University of Toronto, Canada*
K. Shenai
*University of Illinois at Chicago, IL*
M. Shibib
*Lucent Technologies, PA*


IEEE

E D S

# 22nd INTERNATIONAL CONFERENCE ON MICROELECTRONICS

# CALL FOR PAPERS

**Niš, Yugoslavia**
**19-22 September 1999**

organized by
**Yugoslavia IEEE Section - ED/SSC Chapter**
in co-operation with
**Faculty of Electronic Engineering, University of Niš**
**Ei - Holding Co. - Niš**
under the co-sponsorship of
**IEEE Electron Devices Society**
under the auspices of
**Serbian Ministry of Science and Technology**
**Yugoslav Ministry of Science and Development**

The International Conference on Microelectronics (MIEL) is one of the most outstanding European conferences, providing an international forum for the presentation and discussion of the recent developments and future trends in the field of microelectronics. As in previous years, topics covered by the technical programme will include all important aspects of microelectronic devices, circuits and systems, ranging from materials and processes, technologies and devices, device physics and modeling, process and device simulation, circuit design and testing, system design and packaging, and characterization and reliability.

Based on the past decade history, it is expected that the technical programme will consist of about 150 contributed papers by the authors from more than 30 countries all round the world, which will be structured into oral and poster sessions. These papers, together with 15 invited papers, which are to be presented by the world-leading authorities from the field of microelectronics, will form the solid foundation of the Conference MIEL '99. Three related scientific events, namely the workshops "Power Devices and ICs" and "Microsystem Technologies", containing 5 invited papers each, and short course "Low Power Electronics" by Krishna Shenai, will round off the technical programme.

The invited speakers are: Vijay Arora (USA), James Clemens (USA), Bernard Courtois (France), Sorin Cristoloveanu (France), Helmut Detter (Austria), Andrzej Dziedzic (Poland), Wolfgang Ehrfeld (Germany), Fausto Fantini (Italy), Daniel Fleetwood (USA), Masaharu Imai (Japan), Hiroshi Iwai (Japan), Erhard Kohn (Germany), Nikolai Koroteev (Russia), James Kuo (Taiwan), Valentino Liberali (Italy), Juin Liou (USA), Mikael Ostling (Sweden), Pierre Rossel (France), Siegfried Selberherr (Austria), Krishna Shenai (USA), Ayman Shibib (USA), Shi-Chung Sun (Taiwan), Danelle Tanner (USA), Yoshiyuki Uchida (Japan), Paul Zavracky (USA).

The authors of the contributed papers are asked to submit two-page extended abstract (including figures, tables, and references) which will serve as the basis for the papers selection. A cover page of the abstract should include the complete address (including fax and e-mail) of the author to be contacted, as well as the preference for an oral or a poster presentation. The deadline for the receipt of one original plus five copies of the abstract, which should be sent to the Conference Chairman, is the **30 January 1999.**


For further information please contact: *Prof. Dr. N. Stojadinović, MIEL'99 Conference Chairman,* Faculty of Electronic Engineering, University of Niš, Beogradska 14, 18000 Niš, Yugoslavia, Tel: +381 18 529-326, Fax: +381 18 46-180, E-mail: nino@unitop.elfak.ni.ac.yu,
The MIEL'99 Conference website is : http://unitop.elfak.ni.ac.yu/miel/welcome.html

**FROM ELSEVIER SCIENCE...**
SCIENCE PUBLISHER TO THE WORLD

CONTENTS DIRECT

**A FREE alerting service by E-mail for Elsevier Science journals.**

**ContentsDirect** allows you unrivalled access to the tables of contents pages of Elsevier Science journals in the following subject areas:

- Chemistry and Chemical Engineering
- Clinical Medicine
- Computer Science
- Earth and Planetary Sciences
- Economics, Business and Management Science
- Engineering, Energy and Technology
- Environmental Science and Technology
- Life Sciences
- Materials Science
- Mathematics
- Physics and Astronomy
- Social Sciences
- Multidiscipline

**What does ContentsDirect provide?**

- Journal title
- Volume and issue number
- Title of paper
- Names of authors
- Page details
- Anticipated publication date
- News and offers

**Why register to ContentsDirect?**

- Provides via E-mail advance notice of forthcoming papers, allowing you to reserve the issue at your library.
- Provides an invaluable information resource to scientists and researchers requiring the very latest information on soon-to-be published papers.
- Keeps you at the forefront of current research.
- Enables you to create a personal archive of key journal contents in your field of interest.

**How to register for ContentsDirect:**

- The quickest way to register for **ContentsDirect** is via Elsevier Science home page on the world wide web.
- Registration is simple and the selection of titles is provided by access to drop down menus of all subject classifications with a complete listing of all titles available.
- If you wish to select new titles or amend your existing selection - simply re-visit the web site and personally make your changes.

ContentsDirect

ContentsDirect

**Register Now!**

**http://www.elsevier.nl/locate/ContentsDirect**

**NB** If you can't access the web - please note that you can still register by sending an E-mail to **cdsubs@elsevier.co.uk**

ELSEVIER SCIENCE

# Receive Regular News of Elsevier's Publications

Elsevier Science mails information on new and existing publications regularly. If you would like to be added to the mailing list please send us your name and full mailing address, indicating your fields of interest:

a subject areas

- Chemistry and Chemical Engineering
- Clinical Medicine
- Computer Science
- Earth and Planetary Sciences
- Economics, Business and Management Science
- Engineering, Energy and Technology
- Environmental Science and Technology
- Life Sciences
- Materials Science
- Mathematics
- Physics and Astronomy
- Social Sciences


ELSEVIER


PERGAMON


NORTH HOLLAND


EXCERPTA MEDICA

## Regional Sales Offices


**For customers in Europe, Middle East and Africa**
Elsevier Science
Regional Sales Office
Customer Support Department
P.O. Box 211
1000 AE Amsterdam
The Netherlands
Tel: (+31) 20 485 3757
Fax: (+31) 20 485 3432
E-mail: nlinfo-f@elsevier.nl

**For customers in the United States & Canada**
Elsevier Science
Regional Sales Office
Customer Support Department
P.O. Box 945
New York, N.Y. 10159-0945
U.S.A.
Tel: (+1) 212 633 3730
Toll Free number for North-American customers: 1-888-4ES-INFO (437-4636)
Fax: (+1) 212 633 3680
E-mail: usinfo-f@elsevier.com

**For customers in Japan**
Elsevier Science
Regional Sales Office
Customer Support Department
9-15 Higashi-Azabu 1-chome
Minato-ku
Tokyo
106-0044 Japan
Tel: (+81) 3 5561 5033
Fax: (+81) 3 5561 5047
E-mail: info@elsevier.co.jp

**For customers in Asia and Australasia**
Elsevier Science
Regional Sales Office
Customer Support Department
No. 1 Temasek Avenue
#17-01 Millenia Tower
Singapore 039192
Tel: (+65) 434 3727
Fax: (+65) 337 2230
E-mail: asiainfo@elsevier.com.sg

**For customers in Latin America**
Elsevier Science
Regional Sales Office
Rua Sete de Setembro 111/16 Andar
20050-002 Centro
Rio de Janeiro - RJ
Brazil
Tel: (+55) 21 509 5340
Fax: (+55) 21 507 1991
E-mail: elsevier@campus.com.br

# BACK ISSUES OF THIS JOURNAL

*Back issues of this and all other Elsevier Science journals are available in hard copy. New subscribers to a journal may purchase back issues of that publication in hard copy edition at 25% discount off the standard price. Elsevier Science maintains stocks of back issues and orders may be placed with confidence at your nearest Elsevier Science office. Should any issue of a volume be temporarily out of stock at the time of ordering, a high quality photoduplicated copy will be supplied at no extra charge to complete your order.*

## SAVE UP TO 25% BY PURCHASING COMPLETE SETS

Customers wishing to purchase complete sets can do so at a saving of 25% off the individual volume price.

## MICROFORM EDITIONS

Back issues in microform of Elsevier Science research journals are also available. For further information please apply to your nearest Elsevier Science office.

## BACK ISSUES PRICE LIST

Full details of the rates of back issues of all Elsevier Science journals can be found in our Back Issues Price List. Please contact your nearest Elsevier Science office for a copy.

## Regional Sales Offices

**For customers in Europe, Middle East and Africa**
Elsevier Science, Regional Sales Office, Customer Support Department, P.O. Box 211, 1000 AE Amsterdam, The Netherlands
Tel: (+31) 20 485 3757, Fax: (+31) 20 485 3432, E-mail: nlinfo-f@elsevier.nl

**For customers in the United States & Canada**
Elsevier Science, Regional Sales Office, Customer Support Department, P.O. Box 945, New York, N.Y. 10159-0945, U.S.A.
Tel: (+1) 212 633 3730, Toll Free number for North-American customers: 1-888-4ES-INFO (437-4636), Fax: (+1) 212 633 3680
E-mail: usinfo-f@elsevier.com

**For customers in Japan**
Elsevier Science, Regional Sales Office, Customer Support Department, 9-15 Higashi-Azabu, 1-chome, Minato-ku, Tokyo, 106-0044 Japan
Tel: (+81) 3 5561 5033, Fax: (+81) 3 5561 5047, E-mail: info@elsevier.co.jp

**For customers in Asia and Australasia**
Elsevier Science, Regional Sales Office, Customer Support Department, No. 1 Temasek Avenue, #17-01 Millenia Tower, Singapore 039192
Tel: (+65) 434 3727, Fax: (+65) 337 2230, E-mail: asiainfo@elsevier.com.sg

**For customers in Latin America**
Elsevier Science, Regional Sales Office, Rua Sete de Setembro 111/16 Andar, 20050-002 Centro, Rio de Janeiro - RJ, Brazil
Tel: (+55) 21 509 5340, Fax: (+55) 21 507 1991, E-mail: elsevier@campus.com.br


ELSEVIER


PERGAMON



NORTH HOLLAND



EXCERPTA MEDICA

AVAILABLE AT YOUR FINGERTIPS...

# THE ELSEVIER SCIENCE CATALOGUE ON THE INTERNET

**Features include:**

ELSEVIER SCIENCE

*Catalogue on* INTERNET

- ORDER YOUR **FREE** SAMPLE COPY OF *JOURNAL/BOOK TITLE* TODAY BY CHECKING ITS CATALOGUE ENTRY

**Other features include:-**

- All the journals, with complete information about journal editors and editorial boards
- Listings of special issues and volumes
- Listings of recently published papers for many journals
- Complete descriptions and contents lists of book titles
- Clippings of independent reviews of published books
- Book series, dictionaries, reference works
- Electronic and CD-ROM products
- Full text search facilities
- Ordering facilities
- Print options
- Hypertext links
- Alerting facility for new & forthcoming publications
- Updated monthly

ELSEVIER SCIENCE

*Catalogue on* INTERNET

**TRY IT TODAY!**

http://www.elsevier.nl/

For further queries please contact:-

Customer Support Department
Tel: +1 (212) 633 3730
Fax: +1 (212) 633 3680
E-mail: usinfo-f@elsevier.com

Customer Support Department
Tel: +31 (20) 485 3757
Fax: +31 (20) 485 3432
E-mail: nlinfo-f@elsevier.nl

CO7A05/p15

ELSEVIER | PERGAMON | NORTH HOLLAND | EXCERPTA MEDICA

# JOURNAL PACKAGES

Money-saving ways
to buy journals such as...

Atmospheric Environment
Biochemical Pharmacology
Environmental Pollution
Life Sciences
Neuropharmacology
Neuroscience
Tetrahedron Letters
Trends in Pharmacological Sciences

## Which libraries benefit?

*Those libraries which take all the journals from a package will see an immediate benefit and saving. Libraries which take all but one of the journals may find that they can buy the missing title for a nominal amount.*

## Take advantage of these packages

*Subscription Agents have been informed about these money-saving packages and are in a position to help librarians take advantage of the savings offered. Contact your nearest Regional Sales Office directly for more information and help.*


ELSEVIER


PERGAMON



NORTH HOLLAND



EXCERPTA MEDICA

## Regional Sales Offices

**For customers in Europe, Middle East and Africa**
Elsevier Science
Regional Sales Office
Customer Support Department
P.O. Box 211
1000 AE Amsterdam
The Netherlands
Tel: (+31) 20 485 3757
Fax: (+31) 20 485 3432
E-mail: nlinfo-f@elsevier.nl

**For customers in the United States & Canada**
Elsevier Science
Regional Sales Office
Customer Support Department
P.O. Box 945
New York, N.Y. 10159-0945
U.S.A.
Tel: (+1) 212 633 3730
Toll Free number for North-American customers:
1-888-4ES-INFO (437-4636)
Fax: (+1) 212 633 3680
E-mail: usinfo-f@elsevier.com

**For customers in Japan**
Elsevier Science
Regional Sales Office
Customer Support Department
9-15 Higashi-Azabu
1-chome
Minato-ku
Tokyo
106-0044 Japan
Tel: (+81) 3 5561 5033
Fax: (+81) 3 5561 5047
E-mail: info@elsevier.co.jp

**For customers in Asia and Australasia**
Elsevier Science
Regional Sales Office
Customer Support Department
No. 1 Temasek Avenue
#17-01 Millenia Tower
Singapore 039192
Tel: (+65) 434 3727
Fax: (+65) 337 2230
E-mail: asiainfo@elsevier.com.sg

**For customers in Latin America**
Elsevier Science
Regional Sales Office
Rua Sete de Setembro 111/16
Andar
20050-002 Centro
Rio de Janeiro - RJ
Brazil
Tel: (+55) 21 509 5340
Fax: (+55) 21 507 1991
E-mail: elsevier@campus.com.br